Machine Learning–Driven Frameworks for Predicting Stroke Risk Factors: A Comprehensive Systematic Review

Dr. Ahmad Prasetyo
Department of Applied Sciences and Technology
Jakarta Institute of Advanced Scientific Research
Jakarta, Indonesia

Dr. Rina Kusumawati
Faculty of Applied Research and Innovation
Bandung National University of Science and Technology
Bandung, Indonesia

Received: 28 Mar 2025 | Received Revised Version: 18 Apr 2025 | Accepted: 29 May 2026 | Published: 01 June 2026

Volume 08 Issue 06 2026 |

ABSTRACT

Stroke remains one of the leading causes of mortality and long-term disability worldwide, creating substantial clinical, economic, and social burdens. Despite significant advances in prevention and treatment, the increasing prevalence of cardiovascular diseases, aging populations, and lifestyle-related risk factors continue to elevate stroke incidence across diverse populations. Traditional stroke risk assessment approaches rely primarily on statistical models and clinician-based evaluations; however, these methods often struggle to capture complex nonlinear interactions among multiple risk determinants. The emergence of machine learning (ML) has introduced innovative opportunities for improving predictive accuracy, enabling personalized risk assessment, and supporting proactive healthcare interventions. This systematic review critically examines the role of machine learning–driven frameworks in predicting stroke risk factors, synthesizing evidence from contemporary studies addressing epidemiological trends, clinical determinants, predictive analytics, and computational modeling approaches. The review analyzes existing machine learning methodologies, including nonlinear predictive systems, cardiovascular risk-based frameworks, and healthcare analytics models applied to stroke prediction. Furthermore, it evaluates the strengths, limitations, and practical implications of ML-based risk prediction systems within clinical environments. Findings indicate that machine learning models consistently outperform conventional risk stratification approaches in handling multidimensional datasets, identifying hidden relationships among variables, and supporting individualized prediction. However, challenges related to data quality, model interpretability, generalizability, and clinical integration remain significant barriers to widespread adoption. The review proposes a conceptual framework integrating epidemiological, clinical, and computational dimensions to strengthen future stroke prediction systems. The study contributes to the growing literature on intelligent healthcare analytics by providing a comprehensive assessment of machine learning applications for stroke risk prediction and identifying future directions for research and implementation.

Keywords: Stroke Prediction, Machine Learning, Risk Factors, Predictive Analytics, Artificial Intelligence in Healthcare, Clinical Decision Support, Cardiovascular Risk Assessment, Healthcare Informatics, Systematic Review, Disease Prevention.

Cite This Article: Prasetyo, D. A., & Kusumawati, D. R. (2026). Machine Learning–Driven Frameworks for Predicting Stroke Risk Factors: A Comprehensive Systematic Review. The American Journal of Applied Sciences, 8(06), 1–12. Retrieved from https://www.theamericanjournals.com/index.php/tajas/article/view/8022

1.Introduction

Background

Stroke represents a major public health challenge worldwide and remains among the foremost causes of death and disability. According to global stroke reports, millions of individuals experience stroke annually, resulting in substantial healthcare expenditures and long-term socioeconomic consequences (Feigin et al., 2022). The increasing burden of stroke is influenced by demographic transitions, population aging, urbanization, lifestyle changes, and the growing prevalence of cardiovascular risk factors such as hypertension, diabetes mellitus, obesity, and dyslipidemia (Thayabaranathan et al., 2022).

Clinically, stroke is characterized by an abrupt interruption of cerebral blood flow resulting in neurological dysfunction. Ischemic stroke constitutes the majority of stroke cases and often emerges from thrombotic or embolic vascular occlusions, whereas hemorrhagic stroke results from intracerebral or subarachnoid bleeding (Murphy and Werring, 2020). The multifactorial nature of stroke development complicates early identification of high-risk individuals and creates challenges for preventive healthcare systems.

Historically, stroke risk assessment has relied on epidemiological observations and statistical prediction tools. Conventional methods, while valuable, often assume linear relationships among variables and may inadequately account for the dynamic interactions among biological, environmental, and behavioral determinants. Advances in digital health technologies and electronic medical records have generated unprecedented volumes of healthcare data, creating opportunities for sophisticated predictive analytics and machine learning applications (Deo, 2015).

Machine learning has emerged as a transformative technology capable of identifying complex patterns within large datasets. Unlike traditional statistical approaches, machine learning algorithms can process nonlinear relationships, adapt to multidimensional variables, and continuously improve predictive performance through iterative learning processes (Ibrahim and Saber, 2023). These characteristics have positioned machine learning as a promising solution for improving stroke risk prediction and supporting precision medicine initiatives.

The increasing integration of artificial intelligence into healthcare has accelerated research efforts focused on stroke prediction. Recent investigations have demonstrated that machine learning algorithms can effectively identify individuals at elevated stroke risk by analyzing demographic characteristics, cardiovascular indicators, clinical histories, laboratory measurements, and lifestyle variables (Orfanoudaki et al., 2020). These developments suggest substantial potential for enhancing preventive healthcare strategies and reducing stroke-related morbidity and mortality.

Problem Statement

Despite significant progress in understanding stroke pathophysiology and risk determinants, accurate prediction of stroke occurrence remains a persistent challenge. Existing risk prediction models frequently demonstrate limitations related to population specificity, restricted variable inclusion, and inadequate representation of nonlinear interactions among risk factors. Furthermore, many traditional frameworks fail to fully exploit the growing availability of healthcare data generated through electronic health systems, wearable technologies, and population health databases.

Although machine learning technologies offer considerable promise, evidence regarding their effectiveness, implementation challenges, and comparative advantages remains fragmented across multiple studies. Variability in datasets, algorithm selection, validation methodologies, and performance metrics has created uncertainty regarding optimal predictive frameworks. Consequently, a comprehensive synthesis of current evidence is necessary to clarify the role of machine learning in stroke risk prediction.

Research Objectives

This systematic review aims to:

1. Examine major stroke risk factors identified in contemporary literature.
2. Evaluate the application of machine learning techniques in stroke risk prediction.
3. Analyze existing predictive frameworks and computational models.
4. Identify strengths, limitations, and implementation challenges associated with machine learning approaches.
5. Propose future research directions for

improving machine learning–driven stroke prediction systems.

Scope and Significance

The study focuses exclusively on literature addressing stroke risk factors, epidemiological determinants, predictive analytics, and machine learning applications relevant to stroke prediction. By synthesizing findings from clinical, epidemiological, and computational perspectives, the review contributes to a deeper understanding of how machine learning can support preventive healthcare strategies.

The significance of this review extends beyond academic inquiry. Accurate prediction of stroke risk enables earlier interventions, improved allocation of healthcare resources, personalized treatment planning, and enhanced patient outcomes. As healthcare systems increasingly adopt data-driven decision-making processes, understanding the capabilities and limitations of machine learning becomes essential for clinicians, researchers, policymakers, and healthcare administrators.

2. Literature Review

Global Burden and Epidemiology of Stroke

The growing global burden of stroke has stimulated extensive research into prevention and risk assessment strategies. According to the World Stroke Organization, stroke continues to rank among the leading causes of mortality and disability worldwide, affecting populations across developed and developing countries (Feigin et al., 2022). The epidemiological landscape reveals increasing incidence rates in regions experiencing rapid demographic and lifestyle transitions.

Global statistics indicate substantial variations in stroke prevalence, mortality, and disability-adjusted life years across geographic regions (Thayabaranathan et al., 2022). These differences reflect disparities in healthcare access, socioeconomic conditions, public health infrastructure, and exposure to modifiable risk factors. The persistence of these disparities highlights the necessity of targeted prevention strategies supported by robust predictive frameworks.

Forecasting analyses further suggest that stroke incidence will continue to rise in coming decades unless preventive interventions become more effective and widely implemented (Ovbiagele et al., 2013). Consequently, early identification of high-risk individuals has become a central objective within contemporary stroke prevention research.

Clinical and Biological Risk Factors

The literature consistently identifies multiple clinical and biological determinants associated with stroke occurrence. Hypertension remains the most influential modifiable risk factor, contributing significantly to both ischemic and hemorrhagic stroke development (Sacco, 1997). Additional risk factors include diabetes mellitus, hyperlipidemia, obesity, smoking, atrial fibrillation, sedentary lifestyle, and cardiovascular diseases.

Stroke pathophysiology involves complex interactions among vascular, inflammatory, metabolic, and neurological mechanisms. Murphy and Werring (2020) emphasize that stroke development cannot be attributed to a single determinant; rather, it results from interconnected biological processes affecting cerebral circulation and vascular integrity.

Neuroinflammatory mechanisms have gained increasing attention within stroke research. Maida et al. (2020) demonstrate that inflammatory pathways contribute significantly to ischemic stroke progression and outcomes. Their findings suggest that inflammatory biomarkers may serve as valuable predictors within future machine learning models designed for stroke risk assessment.

Cardioembolic conditions represent another major contributor to stroke incidence. Cardiovascular abnormalities frequently interact with metabolic and inflammatory factors, creating multidimensional risk profiles that challenge conventional predictive methodologies. These observations support the need for advanced computational approaches capable of managing complex variable interactions.

Demographic and Environmental Determinants

Demographic characteristics significantly influence stroke risk. Age remains one of the strongest predictors, with risk increasing substantially among older populations (Tsao et al., 2022). Gender differences have also been observed, although findings vary across populations and healthcare settings.

Environmental and socioeconomic factors further complicate risk assessment. Nakibuuka et al. (2015) found notable differences in stroke risk factors between rural and urban communities. Urban populations often exhibit higher prevalence of lifestyle-related risk factors, whereas rural populations may experience barriers to preventive healthcare services.

Infectious diseases have emerged as additional considerations within stroke prediction research. South et al. (2020) revisited the relationship between preceding infections and stroke risk, highlighting how inflammatory responses associated with infectious conditions may contribute to cerebrovascular events. The COVID-19 pandemic further reinforced the importance of considering infection-related variables within predictive models.

Traditional Stroke Risk Prediction Models

Conventional stroke prediction frameworks have historically relied on statistical methodologies and epidemiological observations. These models typically incorporate demographic, clinical, and behavioral variables to estimate future stroke risk. While valuable for population-level assessment, traditional approaches often assume linear relationships among predictors.

The Framingham Stroke Risk Profile represents one of the most influential risk assessment tools developed for stroke prediction. However, limitations associated with linear modeling assumptions have prompted researchers to investigate alternative computational strategies (Orfanoudaki et al., 2020).

Traditional models generally demonstrate acceptable predictive performance within specific populations but may experience reduced accuracy when applied across diverse demographic groups. Additionally, these frameworks frequently struggle to accommodate high-dimensional datasets generated through modern healthcare technologies.

Emergence of Machine Learning in Healthcare

Machine learning has transformed healthcare analytics by enabling automated pattern recognition and predictive modeling across large and complex datasets. Deo (2015) describes machine learning as a paradigm shift in medical decision-making, allowing computational systems to identify relationships that may remain undetected through conventional analytical methods.

The application of machine learning extends across disease diagnosis, prognosis, treatment optimization, and preventive healthcare. Ibrahim and Saber (2023) emphasize that predictive analytics technologies have become increasingly important for advancing disease prevention initiatives and improving healthcare outcomes.

Machine learning algorithms possess several advantages relevant to stroke prediction. These include the ability to process nonlinear relationships, manage multidimensional variables, adapt to evolving datasets, and generate individualized risk assessments. Such capabilities align closely with the multifactorial nature of stroke development.

Machine Learning Applications in Stroke Risk Prediction

Recent studies demonstrate growing interest in machine learning–based stroke prediction systems. One of the most significant contributions is the development of cardiovascular risk factor–based prediction models capable of identifying high-risk individuals before clinical manifestations occur. The Suita Study developed a stroke risk prediction model using cardiovascular risk indicators and demonstrated strong predictive potential across population-based cohorts (Arafa et al., 2022).

Importantly, the findings of Arafa et al. (2022) indicate that integrating multiple cardiovascular variables substantially improves predictive performance compared with isolated risk assessments. The study provides evidence that machine learning-compatible frameworks can enhance risk stratification accuracy while supporting personalized preventive interventions.

Further evidence supporting advanced computational approaches emerges from Orfanoudaki et al. (2020), who demonstrated that stroke risk exhibits significant nonlinear characteristics. Their analysis showed that machine learning methods captured complex relationships overlooked by conventional linear models, leading to improved predictive accuracy.

The relevance of the Suita Study is particularly noteworthy because it illustrates how population-based cardiovascular data can be translated into practical predictive systems (Arafa et al., 2022). Moreover, the study highlights the growing convergence between epidemiological research and artificial intelligence-driven healthcare analytics.

The literature also suggests that future predictive frameworks should incorporate broader variable categories, including inflammatory markers, socioeconomic indicators, behavioral characteristics, and environmental exposures. Such integration would enhance model comprehensiveness and improve individualized prediction capabilities (Arafa et al., 2022).

Research Gaps and Theoretical Positioning

Despite encouraging progress, several research gaps remain evident. First, many machine learning studies rely on relatively homogeneous datasets, limiting generalizability across populations. Second, variability in algorithm selection and validation procedures complicates comparisons among studies. Third, model interpretability remains a significant challenge, particularly within clinical environments where transparent decision-making is essential.

Another critical gap involves the integration of epidemiological, biological, and computational perspectives into unified predictive frameworks. Existing research frequently focuses on algorithm performance without adequately addressing clinical implementation considerations.

Theoretically, this review positions machine learning–driven stroke prediction within a multidimensional framework combining cardiovascular epidemiology, predictive analytics, and precision medicine. Such positioning recognizes that effective stroke prediction requires both accurate computational modeling and meaningful clinical applicability.

3. Methodology

Research Design

This study adopts a systematic review methodology to critically evaluate machine learning–driven frameworks for stroke risk factor prediction. The review synthesizes evidence exclusively from the provided scholarly sources and integrates epidemiological, clinical, and computational perspectives. Unlike a conventional narrative review, the present approach emphasizes analytical comparison of studies, conceptual framework development, and identification of methodological patterns relevant to predictive healthcare systems.

The methodological foundation is based on three interconnected dimensions: stroke epidemiology, machine learning analytics, and clinical risk prediction. The integration of these dimensions allows for a comprehensive assessment of how computational intelligence can improve the identification of individuals at elevated stroke risk.

The review follows four sequential stages. The first stage involves identification and classification of stroke risk factors discussed within the literature. The second stage examines predictive modeling approaches used in healthcare analytics. The third stage evaluates machine learning applications specific to stroke prediction. The fourth stage synthesizes findings into a conceptual framework capable of supporting future research and clinical implementation.

Conceptual Foundation of Stroke Risk Prediction

Stroke prediction is fundamentally a risk stratification process designed to estimate the probability that an individual will experience a cerebrovascular event within a specified time period. Traditional prediction systems rely on predefined risk variables and statistical associations derived from population studies. While these approaches have contributed significantly to preventive medicine, they are constrained by assumptions regarding linearity, variable independence, and limited adaptability.

Machine learning introduces a fundamentally different analytical paradigm. Rather than relying solely on predefined relationships, machine learning algorithms learn patterns directly from data. This capability is particularly relevant for stroke prediction because stroke arises from complex interactions among cardiovascular, metabolic, inflammatory, demographic, and behavioral factors.

The literature indicates that stroke development is influenced by multiple overlapping mechanisms, including vascular dysfunction, neuroinflammation, cardiac abnormalities, and lifestyle-related risk exposures (Maida et al., 2020; Murphy and Werring, 2020). Consequently, effective prediction systems must capture multidimensional relationships rather than isolated variables.

The conceptual foundation adopted in this review assumes that stroke risk prediction is most effective when clinical, epidemiological, and computational information are integrated into a unified analytical framework.

Classification of Stroke Risk Factors

The reviewed literature identifies several categories of stroke risk factors that form the basis of predictive modeling systems.

Demographic Factors

Age remains one of the strongest predictors of stroke incidence. Epidemiological studies consistently demonstrate increased stroke risk among older populations (Tsao et al., 2022; Thayabaranathan et al., 2022). Gender differences have also been observed, although their predictive significance varies across populations.

Geographical location and community characteristics contribute additional predictive information. Research conducted in Uganda demonstrated meaningful differences between urban and rural stroke risk profiles, indicating the importance of environmental context in risk assessment (Nakibuuka et al., 2015).

Cardiovascular Factors

Cardiovascular variables represent the most frequently utilized predictors within stroke risk models. These include:

- Hypertension
- Coronary artery disease
- Atrial fibrillation
- Hyperlipidemia
- Heart failure
- Vascular abnormalities

The Suita Study demonstrated that combining multiple cardiovascular indicators substantially improves predictive performance and provides a robust foundation for stroke risk assessment (Arafa et al., 2022).

Metabolic Factors

Metabolic disorders contribute significantly to stroke susceptibility. Diabetes mellitus, obesity, insulin resistance, and abnormal lipid metabolism are repeatedly identified as major contributors to cerebrovascular disease (Saini and Gurvendra, 2022).

From a machine learning perspective, metabolic indicators are particularly valuable because they can be quantified objectively and monitored continuously through healthcare information systems.

Behavioral and Lifestyle Factors

Lifestyle behaviors influence stroke occurrence through both direct and indirect pathways. Smoking, alcohol consumption, physical inactivity, dietary habits, and medication adherence affect cardiovascular health and consequently influence stroke risk.

Behavioral variables often interact with biological factors, creating nonlinear risk relationships that traditional statistical models may fail to capture effectively.

Inflammatory and Biological Markers

Recent literature highlights the increasing importance of inflammatory pathways in stroke development. Neuroinflammatory mechanisms contribute to vascular damage, plaque instability, and thrombotic events (Maida et al., 2020).

The inclusion of inflammatory biomarkers within machine learning frameworks may enhance predictive performance by providing early indicators of physiological deterioration before clinical symptoms become evident.

Machine Learning Architecture for Stroke Prediction

Based on the reviewed studies, an effective machine learning framework for stroke prediction consists of five interconnected layers.

Layer 1: Data Acquisition

The first layer involves collecting relevant patient information from multiple sources. Typical data categories include:

- Electronic health records
- Laboratory reports
- Cardiovascular assessments
- Medical imaging data
- Lifestyle surveys
- Demographic databases

The quality of prediction depends heavily on the completeness and reliability of input data. Missing values, inconsistent records, and measurement errors can significantly reduce model performance.

Layer 2: Data Preprocessing

Healthcare data frequently contain inconsistencies that require correction before analysis. Preprocessing activities include:

- Missing value management

- Data normalization
- Feature scaling
- Noise reduction
- Variable encoding

Preprocessing transforms raw healthcare information into structured datasets suitable for machine learning algorithms.

Layer 3: Feature Selection

Feature selection identifies variables that contribute most significantly to predictive performance.

Examples include:

- Blood pressure
- Cholesterol levels
- Age
- Diabetes status
- Smoking history
- Previous cardiovascular events

Feature selection improves model efficiency while reducing computational complexity and overfitting risks.

Layer 4: Predictive Modeling

This layer represents the core analytical component of the framework.

Machine learning algorithms learn relationships among variables and generate stroke risk predictions. Unlike conventional statistical models, machine learning systems continuously identify complex interactions among predictors.

The literature suggests that nonlinear analytical approaches are particularly valuable because stroke risk does not develop through strictly linear mechanisms (Orfanoudaki et al., 2020).

Layer 5: Clinical Decision Support

The final layer converts predictive outputs into actionable healthcare information.

Clinical decision support systems may provide:

- Risk scores
- Early warning alerts
- Preventive intervention recommendations
- Patient prioritization guidance

The objective is not merely prediction but practical support for healthcare decision-making.

Categories of Machine Learning Models

Supervised Learning Models

Supervised learning represents the most commonly applied machine learning paradigm in healthcare prediction.

These models learn relationships between predictor variables and known outcomes.

Applications include:

- Stroke occurrence prediction
- Risk stratification
- Outcome forecasting

The Suita Study exemplifies supervised learning principles by using cardiovascular risk factors to predict future stroke events (Arafa et al., 2022).

Advantages include:

- High predictive accuracy
- Clear evaluation metrics
- Strong clinical applicability

Limitations include dependence on high-quality labeled datasets.

Unsupervised Learning Models

Unsupervised learning identifies hidden structures within datasets without predefined outcome labels.

Potential applications include:

- Patient clustering
- Risk subgroup identification
- Population segmentation

Although less common in stroke prediction, unsupervised techniques may reveal previously unrecognized patterns among risk factors.

Hybrid Learning Frameworks

Hybrid systems combine multiple analytical approaches to improve performance.

For example:

- Epidemiological models may identify important risk factors.
- Machine learning models may analyze interactions among variables.
- Clinical systems may translate predictions into interventions.

Hybrid approaches align closely with precision medicine principles and offer promising directions for future research.

Proposed Integrated Stroke Risk Prediction Framework

Based on evidence synthesized from the literature, this review proposes an Integrated Machine Learning Stroke Risk Prediction Framework consisting of four major domains.

Domain 1: Population Health Intelligence

This domain incorporates epidemiological indicators such as:

- Age distributions
- Geographic factors
- Socioeconomic conditions
- Population disease prevalence

Global stroke reports emphasize the importance of considering population-level determinants when designing predictive systems (Feigin et al., 2022).

Domain 2: Clinical Risk Intelligence

Clinical risk intelligence includes:

- Cardiovascular measurements
- Neurological assessments
- Laboratory indicators
- Comorbidity profiles

The framework recognizes that stroke prediction must remain grounded in clinical evidence.

Domain 3: Machine Learning Analytics Engine

This domain performs:

- Pattern recognition
- Risk scoring
- Feature interaction analysis
- Predictive classification

Evidence from nonlinear risk modeling studies suggests that machine learning engines can identify relationships overlooked by traditional methodologies (Orfanoudaki et al., 2020).

Domain 4: Preventive Decision Support

The final domain translates predictions into actionable recommendations.

Examples include:

- Lifestyle interventions
- Medication optimization
- Monitoring protocols
- Preventive screening programs

The ultimate objective is reducing stroke incidence through early intervention.

Validation Considerations

Reliable predictive systems require rigorous validation procedures.

Three forms of validation are particularly important:

Internal Validation

Internal validation assesses performance within the original dataset.

Benefits include:

- Early model assessment
- Detection of overfitting
- Performance optimization

External Validation

External validation evaluates performance using independent populations.

This step is essential because stroke risk factors vary across geographic, demographic, and socioeconomic contexts.

The differences observed between rural and urban populations demonstrate the necessity of testing predictive systems across diverse settings (Nakibuuka et al., 2015).

Clinical Validation

Clinical validation examines whether predictive systems improve healthcare outcomes in real-world environments.

Prediction accuracy alone does not guarantee clinical usefulness. Models must also support effective decision-making and patient management.

Ethical and Implementation Considerations

Machine learning applications in healthcare introduce important ethical considerations.

Data Privacy

Healthcare datasets contain highly sensitive patient information. Effective governance mechanisms are necessary to protect confidentiality and ensure regulatory compliance.

Algorithmic Bias

Models trained on nonrepresentative populations may produce biased predictions.

For example, risk factors identified within one demographic group may not generalize effectively to another population.

Explainability

Healthcare professionals require transparent explanations for predictive outcomes.

Complex machine learning systems may achieve high accuracy but encounter resistance if clinicians cannot understand their reasoning processes.

Clinical Integration

Successful implementation depends on compatibility with existing healthcare workflows.

Prediction systems should complement clinical expertise rather than replace physician judgment.

Methodological Synthesis

The methodological analysis demonstrates that machine learning frameworks offer substantial advantages for stroke risk prediction by integrating multidimensional risk factors, identifying nonlinear relationships, and supporting individualized assessment. The reviewed evidence indicates that cardiovascular variables remain the strongest predictors of stroke occurrence, while emerging research highlights the importance of incorporating inflammatory, behavioral, and environmental determinants. The repeated success of cardiovascular-based predictive systems, particularly the model developed by Arafa et al. (2022), suggests that future frameworks should combine established clinical indicators with advanced machine learning analytics to maximize predictive effectiveness.

4. Results

The systematic review reveals that machine learning–driven frameworks have substantially improved the ability to predict stroke risk compared with traditional epidemiological and statistical approaches. Across the reviewed literature, stroke prediction accuracy is strengthened when multiple categories of risk factors are integrated into a single analytical model rather than assessed independently.

A consistent finding is the dominant influence of cardiovascular risk factors on stroke occurrence. Hypertension, atrial fibrillation, coronary artery disease, diabetes mellitus, and lipid abnormalities emerged as the most frequently reported predictors of future stroke events (Sacco, 1997; Tsao et al., 2022; Tsao et al., 2023). These variables form the foundation of most predictive systems and remain critical regardless of the computational method employed.

The review further indicates that stroke risk prediction benefits from incorporating demographic and environmental variables. Age consistently demonstrates strong predictive value, while geographical and community-level characteristics contribute additional explanatory power (Nakibuuka et al., 2015). Such findings suggest that stroke prediction should extend beyond purely clinical indicators and include contextual determinants of health.

Evidence from machine learning studies demonstrates superior capability in capturing nonlinear relationships among risk factors. Orfanoudaki et al. (2020) showed that stroke risk progression does not follow strictly linear patterns, thereby limiting the effectiveness of conventional risk equations. Machine learning algorithms overcome this limitation by identifying hidden interactions among variables and generating more individualized predictions.

A particularly important finding concerns the predictive value of integrated cardiovascular models. The Suita Study developed a stroke risk prediction model based on cardiovascular risk factors and demonstrated strong predictive performance within population-based cohorts (Arafa et al., 2022). The study provides empirical evidence that combining multiple cardiovascular indicators improves risk stratification and supports preventive intervention planning. The significance of this finding is reinforced by the fact that cardiovascular variables remain measurable, clinically relevant, and routinely available within healthcare systems.

Another emerging trend involves the incorporation of biological and inflammatory markers. Neuroinflammatory mechanisms have been increasingly associated with stroke development and progression (Maida et al., 2020). The literature suggests that future machine learning frameworks may achieve higher predictive accuracy through integration of inflammatory biomarkers alongside traditional cardiovascular indicators.

The review also highlights the growing role of predictive analytics in healthcare decision-making. Machine learning systems not only estimate future stroke risk but also support early disease detection, patient prioritization, and preventive care planning (Ibrahim and Saber, 2023; Rasool et al., 2023). Consequently, predictive frameworks are evolving from analytical tools into components of broader clinical decision-support systems.

Despite these advantages, significant implementation challenges remain. Data quality issues, lack of standardization, model interpretability concerns, and limited external validation continue to restrict widespread clinical adoption. Variability in study designs and population characteristics further complicates comparisons across predictive models.

Overall, the evidence indicates that machine learning frameworks provide meaningful improvements in stroke risk prediction while simultaneously revealing important opportunities for further methodological refinement and clinical integration.

5. Discussion

The findings of this review demonstrate that machine learning represents a significant advancement in stroke risk prediction research. Traditional risk assessment models have historically provided valuable clinical guidance; however, their reliance on linear assumptions limits their capacity to capture the multifactorial nature of stroke development. Machine learning addresses this limitation by identifying complex interactions among demographic, clinical, biological, and behavioral variables.

From a theoretical perspective, the reviewed literature supports the transition from population-based prediction toward individualized risk assessment. Conventional models estimate risk based on average population characteristics, whereas machine learning systems generate personalized predictions derived from multidimensional datasets. This shift aligns with broader developments in precision medicine and data-driven healthcare.

The findings also emphasize the continuing importance of cardiovascular determinants. Although machine learning introduces sophisticated analytical capabilities, model performance remains dependent on clinically meaningful predictors. The repeated success of cardiovascular-based frameworks confirms that predictive accuracy results not only from algorithm selection but also from the quality and relevance of input variables. The model developed by Arafa et al. (2022) provides a strong example of how cardiovascular risk indicators can be effectively translated into predictive systems. Furthermore, the study illustrates how epidemiological evidence and computational analytics can be integrated to improve preventive healthcare strategies.

An important implication of the findings is the recognition of stroke as a multidimensional disease process. Biological mechanisms, inflammatory pathways, lifestyle behaviors, and environmental exposures interact continuously throughout disease progression. Machine learning frameworks are uniquely positioned to analyze these interactions because they are not constrained by traditional assumptions regarding variable independence or linearity. The nonlinear characteristics identified by Orfanoudaki et al. (2020) reinforce the need for advanced computational approaches capable of representing real-world clinical complexity.

The review also highlights practical implications for healthcare systems. Early identification of high-risk individuals can facilitate targeted interventions, improve resource allocation, and reduce the long-term burden associated with stroke-related disability. Predictive frameworks may support clinicians in prioritizing preventive measures, optimizing treatment plans, and monitoring vulnerable populations. These capabilities are particularly relevant given the increasing global burden of stroke reported by Feigin et al. (2022) and Thayabaranathan et al. (2022).

Nevertheless, several limitations warrant consideration. First, many machine learning studies utilize datasets derived from specific populations, potentially limiting generalizability. Second, model transparency remains a significant concern. Clinicians may hesitate to adopt highly complex algorithms when decision-making processes are not easily interpretable. Third, healthcare data frequently contain missing values, inconsistencies, and biases that can affect predictive reliability. Finally, the reviewed studies exhibit methodological heterogeneity, making direct comparisons difficult.

Future research should prioritize external validation across diverse populations, development of explainable artificial intelligence techniques, and integration of emerging biomarkers into predictive models. Additional emphasis should be placed on translating predictive performance into measurable clinical outcomes. The ultimate success of machine learning in stroke prediction will depend not only on algorithmic accuracy but also on practical implementation within healthcare environments.

6. Conclusion

Stroke remains a major global health challenge characterized by substantial mortality, disability, and economic burden. The increasing prevalence of cardiovascular diseases, aging populations, and lifestyle-related risk factors underscores the need for effective preventive strategies and accurate risk assessment systems. This systematic review examined the evolving role of machine learning–driven frameworks in predicting stroke risk factors and synthesized evidence from epidemiological, clinical, and computational perspectives.

The review demonstrates that machine learning offers significant advantages over traditional prediction approaches by enabling analysis of complex, multidimensional, and nonlinear relationships among risk determinants. Cardiovascular variables remain the most influential predictors of stroke occurrence, while demographic, environmental, behavioral, and inflammatory factors provide additional predictive value. Evidence from the reviewed literature, particularly the cardiovascular-based model developed by Arafa et al. (2022), highlights the effectiveness of integrated predictive frameworks for identifying high-risk individuals and supporting preventive interventions.

The findings further indicate that machine learning applications extend beyond prediction alone. These systems increasingly contribute to clinical decision support, disease prevention, and personalized healthcare delivery. However, challenges related to data quality, algorithm transparency, validation, and implementation continue to limit widespread adoption. Addressing these barriers will be essential for realizing the full potential of machine learning within stroke prevention programs.

This review contributes to the growing body of knowledge on intelligent healthcare analytics by providing a comprehensive synthesis of current evidence and proposing an integrated conceptual framework for future development. Future investigations should focus on explainable machine learning models, multicenter validation studies, incorporation of emerging biological markers, and real-world clinical deployment. Through continued interdisciplinary collaboration among clinicians, epidemiologists, and data scientists, machine learning has the potential to transform stroke risk prediction and improve preventive healthcare outcomes on a global scale.

References

1. Arafa, A., Y. Kokubo, H.A. Sheerah, Y. Sakai and E. Watanabe et al., 2022. Developing a stroke risk prediction model using cardiovascular risk factors: The suita study. Cerebrovasc. Dis., 51: 323-330.
2. Deo, R.C., 2015. Machine learning in medicine. Circulation, 132: 1920-1930.
3. Feigin, V.L., M. Brainin, B. Norrving, S. Martins and R.L. Sacco et al., 2022. World Stroke Organization (WSO): Global stroke fact sheet 2022. Int. J. Stroke, 17: 18-29.
4. Ibrahim, M.S. and S. Saber, 2023. Machine learning and predictive analytics: Advancing disease prevention in healthcare. J. Contemp. Healthcare Anal., 7: 53-71.
5. Konduru, S.S.T., A. Ranjan, A. Bollisetty and V. Yadla, 2018. Assessment of risk factors Influencing functional outcomes in cerebral stroke patients using modified Rankin scale. World J. Pharm. Pharm. Sci.,

7: 755-769.
6. Maida, C.D., R.L. Norrito, M. Daidone, A. Tuttolomondo and A. Pinto, 2020. Neuroinflammatory mechanisms in ischemic stroke: Focus on cardioembolic stroke, background, and therapeutic approaches. Int. J. Mol. Sci., Vol. 21. 10.3390/ijms21186454.
7. Murphy, S.J.X. and D.J. Werring, 2020. Stroke: Causes and clinical features. Medicine, 48: 561-566.
8. Nakibuuka, J., M. Sajatovic, J. Nankabirwa, A.J. Furlan and J. Kayima et al., 2015. Stroke-risk factors differ between rural and urban communities: Population survey in Central Uganda. Neuroepidemiology, 44: 156-165.
9. Orfanoudaki, A., E. Chesley, C. Cadisch, B. Stein, A. Nouh, M.J. Alberts and D. Bertsimas, 2020. Machine learning provides evidence that stroke risk is not linear: The non-linear Framingham stroke risk score. PLoS ONE, Vol. 15. 10.1371/journal.pone.0232414.
10. Ovbiagele, B., L.B. Goldstein, R.T. Higashida, V.J. Howard and S.C. Johnston et al., 2013. Forecasting the future of stroke in the United States: A policy statement from the American Heart Association and American Stroke Association. Stroke, 44: 2361-2375.
11. Rasool, S., A. Husnain, A. Saeed, A.Y. Gill and H.K. Hussain, 2023. Harnessing predictive power: Exploring the crucial role of machine learning in early disease detection. JURIHUM: J. Inovasi Humaniora, 1: 302-315.
12. Sacco, R.L., 1997. Risk factors, outcomes, and stroke subtypes for ischemic stroke. Neurology, 49: S39-S44.
13. Saceleanu, V.M., C. Toader, H. Ples, R.A. Covache-Busuioc and H.P. Costin et al., 2023. Integrative approaches in acute ischemic stroke: From symptom recognition to future innovations. Biomedicines, Vol. 11. 10.3390/biomedicines11102617.
14. Saini, N.R. and A.L. Gurvendra, 2022. Stroke-related risk factors: A review. Asian Pac. J. Health Sci., 9: 102-107.
15. South, K., L. McCulloch, B.W. McColl, M.S.V. Elkind, S.M. Allan and C.J. Smith, 2020. Preceding infection and risk of stroke: An old concept revived by the COVID-19 pandemic. Int. J. Stroke, 15: 722-732.
16. Thayabaranathan, T., J. Kim, D.A. Cadilhac, A.G. Thrift and G.A. Donnan et al., 2022. Global stroke statistics 2022. Int. J. Stroke, 17: 946-956.
17. Tsao, C.W., A.W. Aday, Z.I. Almarzooq, A. Alonso and A.Z. Beaton et al., 2022. Heart disease and stroke statistics-2022 update: A report from the American Heart Association. Circulation, 145: e153-e639.
18. Tsao, C.W., A.W. Aday, Z.I. Almarzooq, C.A.M. Anderson and P. Arora et al., 2023. Heart disease and stroke statistics-2023 update: A report from the American heart association. Circulation, 147: e93-e621.