World Journal of Advanced Pharmaceutical and Life Sciences

Journal homepage: https://zealjournals.com/wjapls/

ISSN: 2799-0222 (Online)

(Review Article)

Check for updates

Comprehensive review of logistic regression techniques in predicting health outcomes and trends

Kehinde Josephine Olowe [1,*], Ngozi Linda Edoh [2], Stephane Jean Christophe Zouo [3] and Jeremiah Olamijuwon [4]

[1] *Independent Researcher, Atlanta, Georgia, USA.*
[2] *Osiri University Lincoln Nebraska, USA and Apex Home Care INC.*
[3] *Department of Business Administration, Texas A&M University Commerce, Texas USA.*
[4] *Etihuku Pty Ltd, Midrand, Gauteng, South Africa.*

World Journal of Advanced Pharmaceutical and Life Sciences, 2024, 07(02), 016–026

Publication history: Received on 09 October 2024; revised on 19 November 2024; accepted on 22 November 2024

Article DOI: https://doi.org/10.53346/wjapls.2024.7.2.0039

Abstract

Logistic regression is a powerful statistical method widely used in health research to model and predict the probability of binary and categorical outcomes. This comprehensive review explores the application of logistic regression techniques in predicting health outcomes and trends. The review delves into the theoretical foundations of logistic regression, highlighting its core concepts such as odds ratios, logit transformation, and model interpretation. The use of logistic regression in health outcome prediction, particularly in disease risk assessment, clinical decision-making, and public health studies, is thoroughly examined. This explore various logistic regression models, including binary, multinomial, and ordinal logistic regression, and their roles in analyzing health data. Key applications in predicting diseases such as heart disease, diabetes, and cancer are discussed, emphasizing how logistic regression helps identify risk factors and predict patient outcomes. The review also covers advanced techniques, such as regularization methods (e.g., Lasso and Ridge regression), which help handle high-dimensional health data and improve model accuracy. Additionally, the application of logistic regression in evaluating healthcare interventions, understanding epidemiological trends, and informing public health policies is highlighted. Despite its strengths, logistic regression faces challenges such as data quality issues, overfitting, and model assumptions. The review discusses solutions to address these challenges, including techniques for model validation and diagnostics. Furthermore, the integration of logistic regression with other machine learning approaches, such as ensemble methods, is considered as a means to enhance predictive power and robustness. The review concludes by examining future directions in logistic regression for health outcome prediction, including the use of big data, real-time predictive modeling, and efforts to improve model interpretability. Logistic regression remains a cornerstone technique in health research, offering valuable insights for both clinical and public health applications.

Keywords: Logistic Regression Techniques; Health Outcomes; Predictive Modeling; Comprehensive Review

1. Introduction

Logistic regression is a statistical method widely used in health research to model the relationship between one or more independent variables and a binary or categorical dependent variable (Ekpobimi, 2024). This technique is particularly valuable in medical and public health fields, as it enables researchers to predict the probability of a health outcome occurring, based on various risk factors (Kassem *et al.*, 2022). For instance, logistic regression can be used to estimate the likelihood of a patient developing a disease, such as diabetes or heart disease, based on factors like age, gender, lifestyle, and genetic predispositions. Historically, logistic regression emerged from the work of statisticians such as Ronald A (Arinze *et al.*, 2024). Fisher in the early 20th century, who developed foundational statistical methods for

* Corresponding author: Kehinde Josephine Olowe

health-related data analysis. Over the years, logistic regression models evolved to accommodate more complex datasets and to adjust for confounding variables (Adekoya *et al*., 2024). The method has since become an essential tool in the health research toolkit, owing to its ability to model non-linear relationships between predictor variables and outcomes, and its robustness in dealing with binary outcomes such as disease presence or absence. Today, logistic regression plays a central role in epidemiological studies, clinical trials, and public health research, contributing significantly to evidence-based healthcare decision-making (Adekoya *et al*., 2024; Segun-Falade *et al*., 2024).

The primary objective of this review is to provide a comprehensive understanding of logistic regression techniques, particularly in the context of health research. By exploring the fundamental concepts of logistic regression, this review aims to clarify the mathematical and statistical foundations that underlie this method. In addition, we will delve into the practical applications of logistic regression in predicting health outcomes, including its role in identifying risk factors, stratifying patient populations, and informing public health strategies. Furthermore, the review will examine the advancements and variations in logistic regression models, highlighting their capacity to address diverse health-related issues. This includes exploring extended forms of logistic regression, such as multinomial and ordinal logistic regression, which can handle multiple categories of health outcomes, thereby expanding the applicability of this method. By focusing on these key aspects, the review aims to provide a clear and comprehensive resource for researchers and healthcare professionals looking to apply logistic regression in their work.

This review covers several dimensions of logistic regression as it pertains to health research. We will begin with a detailed examination of the theoretical aspects of logistic regression, including its foundational principles, assumptions, and limitations. The review will then progress to methodological advancements, such as improvements in model selection, handling multicollinearity, and addressing issues like overfitting and bias. Furthermore, we will address the challenges researchers face when using logistic regression models in health studies, such as the need for large, well-collected datasets and the difficulty of interpreting model coefficients in complex, multivariable contexts. In addition to these theoretical and methodological aspects, the review will focus on the practical applications of logistic regression in health-related research. This includes case studies and examples from epidemiology, clinical trials, and health outcomes prediction. By examining the diverse ways in which logistic regression is used in health research, this review aims to illustrate both the versatility and the limitations of this statistical technique (Adeniran *et al*., 2024). Ultimately, the review aims to equip researchers, healthcare practitioners, and public health professionals with the knowledge necessary to effectively apply logistic regression in their respective fields.

2. Theoretical Foundation of Logistic Regression

Logistic regression is a statistical model used to predict the probability of a binary outcome based on one or more predictor variables (Efunniyi *et al*., 2024). It is widely applied in health research to model outcomes such as the presence or absence of a disease, death, or recovery. The simplest form, binary logistic regression, is used when the dependent variable has two possible outcomes (e.g., 1 for disease presence, 0 for disease absence). In contrast, multinomial logistic regression is used when the outcome has more than two categories, such as different disease stages (mild, moderate, severe). At the core of logistic regression is the concept of odds and the odds ratio. The odds of an event occurring is the ratio of the probability that the event happens to the probability that it does not happen (Adeniran *et al*., 2024). For example, in a study predicting the likelihood of developing diabetes, the odds of developing diabetes for a particular risk factor (e.g., high blood pressure) are the ratio of those who develop diabetes to those who do not. The odds ratio (OR) is used to quantify how much the odds of an event change with a one-unit increase in the predictor variable. Logistic regression makes several assumptions, including: (1) the independence of observations, (2) a linear relationship between predictor variables and the logit of the outcome, (3) no multicollinearity among predictors, and (4) a large sample size to ensure the stability of parameter estimates.

Binary logistic regression is the most basic form and is used when the outcome has two categories, such as presence or absence of a disease (Ofoegbu *et al*., 2024). However, in many health outcomes, especially when more than two possible categories exist, multinomial logistic regression is applied. For example, in the prediction of the stages of cancer (early, moderate, severe), multinomial logistic regression would model the odds of being in each stage relative to a reference stage. Ordinal logistic regression is used when the outcome variable has ordered categories, but the distances between the categories are not necessarily equal (Akinsulire *et al*., 2024). For example, when assessing the severity of disease (mild, moderate, severe), ordinal logistic regression can model the relationship between predictor variables and ordered health outcomes. This type of regression accounts for the inherent order in the outcome but does not assume that the difference between categories is uniform. To ensure the quality and reliability of a logistic regression model, it is essential to perform diagnostics and validation. One common method is the Akaike Information Criterion (AIC), which assesses model fit while penalizing for the number of predictors used. Lower AIC values indicate a better-fitting model. Other methods, such as goodness-of-fit tests, like the Hosmer-Lemeshow test, assess how well the model fits the data compared to a baseline model. Cross-validation is another technique used to validate the model. It involves splitting the data into multiple subsets and training the model on some subsets while testing it on others, helping to assess the model's generalizability (Iwuanyanwu *et al.*, 2024). Additionally, ROC (Receiver Operating Characteristic) curves are used to evaluate the discriminative ability of the model. The area under the curve (AUC) quantifies the model's ability to distinguish between different outcome classes. A confusion matrix helps assess the performance by showing the true positive, true negative, false positive, and false negative rates (Segun-Falade *et al.*, 2024). These diagnostic tools ensure that logistic regression models provide reliable, interpretable, and valid predictions for health-related outcomes.

2.1. Applications of Logistic Regression in Health Outcome Prediction

Logistic regression is a powerful statistical tool that has been widely applied in various aspects of health research, including disease prediction, public health studies, clinical decision-making, and healthcare policy planning (Okeke *et al.*, 2024; Bakare *et al.*, 2024). Its ability to model binary outcomes makes it particularly valuable for predicting the likelihood of health events, assessing risks, and informing interventions.

One of the primary applications of logistic regression in health research is in disease prediction and risk assessment (Osundare and Ige, 2024). Logistic regression is often used to predict the likelihood of an individual developing a specific disease based on various risk factors. For example, it has been extensively used in predicting heart disease, diabetes, and cancer. These conditions are influenced by multiple factors such as age, gender, lifestyle, family history, and underlying medical conditions. By incorporating these variables into logistic regression models, researchers and clinicians can estimate the probability of a patient developing these diseases, enabling early intervention and preventive care. For example, in cardiovascular health, logistic regression models can quantify the impact of factors like hypertension, smoking, and cholesterol levels on the likelihood of heart disease. Similarly, in diabetes prediction, logistic regression can assess how variables such as body mass index (BMI), family history, and age contribute to the probability of developing Type 2 diabetes (Achumie *et al.*, 2024). These models help identify significant health determinants, providing valuable insights into the most important factors contributing to disease risk, which can guide public health efforts and individual health management strategies.

In public health and epidemiological studies, logistic regression plays a critical role in understanding disease prevalence and identifying risk factors across different populations. It is particularly useful in assessing how social, behavioral, and environmental determinants influence the distribution of diseases (Alemede *et al.*, 2024). By using logistic regression models, researchers can identify risk factors and predict how these factors contribute to health disparities among various populations. For instance, logistic regression has been employed in global health studies to explore the prevalence of infectious diseases like HIV/AIDS or malaria and chronic conditions such as obesity and asthma. It allows researchers to account for variables like socioeconomic status, access to healthcare, geographic location, and education, which are known to influence health outcomes (Ekpobimi *et al.*, 2024). In some settings, it has helped identify behavioral factors (e.g., smoking, physical inactivity) and environmental factors (e.g., air pollution, access to clean water) that disproportionately affect certain populations. This information is critical for public health planning and intervention strategies aimed at reducing disease burden in underserved communities. Logistic regression also facilitates the analysis of longitudinal health data, enabling the identification of trends and potential causative factors. For example, studies tracking the spread of infectious diseases can utilize logistic regression to predict future outbreaks and determine the effectiveness of vaccination programs or public health campaigns.

Logistic regression plays a significant role in clinical decision-making and assessing the effectiveness of health interventions (Mokogwu *et al.*, 2024). In clinical practice, logistic regression models can be used to predict the outcome of medical treatments or interventions based on patient characteristics. For example, models can predict the likelihood of a patient responding to a particular treatment for conditions such as cancer, cardiovascular disease, or depression. These models provide a quantitative basis for making personalized treatment decisions. By incorporating variables such as patient demographics, comorbidities, and treatment history, clinicians can estimate the probability of success for different interventions, helping them tailor treatment plans to individual patients. Logistic regression can also be used to forecast the likelihood of adverse outcomes such as surgical complications or treatment failure, which is crucial for managing patient expectations and optimizing healthcare delivery. In the field of personalized medicine, logistic regression is particularly valuable for identifying which patients are more likely to benefit from specific therapies, including the use of genetic or molecular data (Ezeafulukwe *et al.*, 2024). This allows for more precise and effective treatment regimens, improving patient outcomes and minimizing unnecessary side effects.

Logistic regression is also a critical tool in healthcare policy and planning. It can help policymakers make informed decisions about resource allocation, disease prevention, and health interventions (Usuemerai *et al.*, 2024). By modeling health trends and forecasting healthcare demands based on demographic and epidemiological data, logistic regression aids in the development of targeted policies and public health programs. For example, logistic regression has been used to predict the future demand for healthcare services in different populations, such as the elderly or those with chronic conditions. This enables health systems to allocate resources efficiently, plan for healthcare infrastructure, and develop strategies for managing the rising burden of non-communicable diseases (NCDs). Additionally, logistic regression models can assess the effectiveness of public health policies, such as vaccination campaigns, smoking cessation programs, or nutrition interventions, by comparing the likelihood of desired outcomes in different policy scenarios. Overall, logistic regression is invaluable in forecasting healthcare needs, assessing the impact of health policies, and planning interventions that can improve health outcomes on a global scale (Kassem *et al*., 2024; Nwaimo *et al*., 2024). By integrating logistic regression with other public health data, governments and international organizations can better prepare for emerging health challenges and allocate resources where they are most needed. Logistic regression remains a cornerstone tool in health research and policy, with applications ranging from disease prediction and risk assessment to clinical decision-making and public health planning (Ajiga *et al*., 2024). As healthcare continues to evolve, the integration of logistic regression with emerging technologies such as big data analytics and machine learning will enhance its utility, further supporting evidence-based practices and the advancement of global health.

2.2. Advanced Logistic Regression Techniques in Health Predictions

Logistic regression is a cornerstone technique in health research, widely used for predicting binary outcomes such as disease occurrence or treatment success (Adewumi *et al*., 2024). However, as health data grows increasingly complex, with more variables, interactions, and temporal aspects, advanced logistic regression techniques have become crucial for refining predictions and improving model accuracy. These techniques include regularization methods, interaction effects, multilevel models, and time-to-event analysis, all of which offer more sophisticated ways to analyze health data and predict outcomes. In high-dimensional datasets, where the number of predictors may exceed the number of observations, regularization methods such as Lasso (Least Absolute Shrinkage and Selection Operator) and Ridge regression are invaluable. These methods help to address issues of overfitting and model complexity by imposing penalties on the regression coefficients, thereby shrinking them towards zero. Regularization techniques enhance the model's ability to generalize, making them especially useful when dealing with large-scale health data, such as genomic datasets, electronic health records, or survey data with many variables. Lasso regression performs both variable selection and regularization by shrinking some coefficients to exactly zero, effectively excluding irrelevant predictors from the model (Bakare *et al*., 2024). This is beneficial in identifying the most important risk factors for health outcomes, as it automatically selects a subset of features. Ridge regression, on the other hand, imposes a penalty on the size of coefficients but does not set them to zero. It is useful when dealing with collinearity among predictors or when the dataset has many correlated variables, a common situation in health studies where multiple interrelated factors (such as demographic, behavioral, and clinical variables) influence outcomes. These regularization methods are essential in high-dimensional health data analysis, such as when dealing with genetic or multi-omics data. They help streamline models, reducing noise while maintaining predictive accuracy.

In health outcome prediction, many risk factors do not operate independently; instead, their combined effects often have a significant impact on health outcomes. Therefore, interaction effects between variables, such as age, gender, and comorbidities, should be carefully modeled. Interaction terms in logistic regression capture the combined effect of two or more predictors on the likelihood of an outcome, allowing for more nuanced predictions (Achumie *et al*., 2024). For example, the effect of smoking on lung cancer risk might be different for people of different ages or genders. By including an interaction term between smoking and age, logistic regression models can reveal how the relationship between smoking and lung cancer risk changes with age, providing more accurate predictions (Segun-Falade *et al*., 2024). Similarly, the interaction between socioeconomic status and access to healthcare might explain health disparities across populations. Including interaction terms helps improve the model’s precision, enabling health professionals and policymakers to tailor interventions more effectively to various subgroups.

Health data are often hierarchical in nature, where individuals are nested within larger groups, such as patients within hospitals, or citizens within communities (Usuemerai *et al*., 2024). Standard logistic regression models, which assume independence between observations, may not account for the clustering of data within groups, potentially leading to biased results. To address this, multilevel or hierarchical models are used, which allow for the inclusion of both individual-level and group-level predictors. In hierarchical models, logistic regression can be extended by incorporating random effects for higher-level clusters. For example, when modeling disease outcomes, random effects can account for the variability between hospitals or communities that may influence the likelihood of disease. This approach is particularly useful in public health studies, where geographic or organizational factors can affect health outcomes. Hierarchical models allow for the analysis of nested data structures, improving model accuracy by acknowledging the dependence between data points within the same group (Iwuanyanwu *et al*., 2024). They also provide more reliable estimates for group-level parameters, such as community-level disease rates or hospital-level performance in health interventions.

In many health studies, the timing of an event is as important as the occurrence of the event itself (Alemede *et al*., 2024). For example, predicting the time until a patient develops heart disease or the survival time after a cancer diagnosis requires time-to-event analysis. Traditional logistic regression models do not directly account for time, so extensions such as survival logistic regression are used to incorporate time components into the analysis. Survival analysis techniques, such as the Cox proportional hazards model, can be combined with logistic regression to model the probability of an event occurring over time, adjusting for risk factors (Ezeafulukwe *et al*., 2024). This approach allows for the prediction of health outcomes like the time until disease relapse, recovery, or death. It is particularly useful in oncology and chronic disease management, where the timing of an event (e.g., tumor recurrence or kidney failure) is critical for treatment planning and patient care. By incorporating time into logistic regression models, researchers can provide more dynamic and clinically relevant predictions. For example, combining survival analysis with logistic regression allows for the prediction of long-term health trends in populations, helping to inform both individual treatment decisions and public health strategies. The ability to model survival times also aids in forecasting the impact of various interventions over time, such as the long-term effects of a vaccination program or the success of a new treatment regimen (Ekpobimi *et al*., 2024; Bakare *et al*., 2024).

Advanced logistic regression techniques have revolutionized health outcome predictions by addressing the complexities of modern health data (Osundare and Ige, 2024). Regularization methods like Lasso and Ridge regression enhance model accuracy by reducing overfitting in high-dimensional datasets. Interaction terms allow for a deeper understanding of how combined risk factors influence health outcomes, while multilevel models account for hierarchical data structures, ensuring more reliable results in public health research. Additionally, time-to-event analysis and survival logistic regression provide vital insights into the timing of health events, facilitating long-term predictions and improving clinical decision-making. As health data becomes increasingly complex, these advanced logistic regression methods will continue to drive innovation in health outcome prediction, offering more precise and actionable insights for researchers, clinicians, and policymakers (Mokogwu *et al*., 2024; Nwaimo *et al*., 2024).

2.3. Challenges and Limitations of Logistic Regression in Health Research

Logistic regression is one of the most widely used statistical methods for analyzing binary outcomes in health research, providing valuable insights into disease prediction, risk factors, and treatment effects (Usuemerai *et al*., 2024). However, despite its widespread application, logistic regression faces several challenges and limitations that can affect the accuracy and reliability of health predictions. These challenges range from issues with data quality and model assumptions to concerns regarding overfitting, interpretability, and ethical considerations. Understanding and addressing these challenges is crucial for improving the effectiveness of logistic regression models in health research.

One of the primary challenges in applying logistic regression to health research is dealing with issues related to data quality. Health datasets often suffer from multicollinearity, where predictor variables are highly correlated with each other (Kassem *et al*., 2024). This can distort the estimated coefficients and inflate standard errors, leading to unreliable results. In health research, where variables such as age, socioeconomic status, and lifestyle factors are often interrelated, multicollinearity can make it difficult to disentangle the independent effects of each predictor. To address this, methods such as regularization (e.g., Lasso or Ridge regression) can be employed to reduce multicollinearity by shrinking or eliminating some predictor variables. Another common issue is missing data, which can occur due to incomplete responses in surveys or lost data in medical records. Missing data can lead to biased estimates if not handled properly. Common approaches for managing missing data include imputation techniques (e.g., multiple imputation or predictive modeling) or using complete-case analysis, but each comes with its own set of assumptions and potential biases. Outliers in the dataset can also affect the logistic regression model. Extreme values, if not properly identified and addressed, can disproportionately influence the model's estimates (Segun-Falade *et al*., 2024). Techniques such as data transformation or robust regression methods are often used to mitigate the effect of outliers. Additionally, logistic regression relies on several key assumptions, including the linearity of the logit (the relationship between predictor variables and the log-odds of the outcome). Violations of this assumption can lead to incorrect inferences. For instance, if a continuous predictor variable has a non-linear relationship with the outcome, the logistic model may not fit the data well. In such cases, non-linear transformations or the inclusion of higher-order terms might be necessary.

Another significant challenge in logistic regression is overfitting, which occurs when a model becomes too complex and captures noise or random fluctuations in the data rather than the true underlying relationship (Ibikunle *et al*., 2024). Overfitting results in a model that performs well on the training data but poorly on new, unseen data. This is particularly problematic in health research, where datasets may contain noise or measurement errors. To avoid overfitting, techniques like cross-validation, regularization, and simplifying models by removing unnecessary predictors are essential. On the other hand, underfitting happens when the model is too simple to capture the complexities of the data. This often leads to poor predictive performance. In health research, underfitting can occur when important predictors are left out of the model, leading to biased estimates of the effects. Ensuring model adequacy through proper feature selection and diagnostic checks is crucial for achieving a balance between overfitting and underfitting.

Logistic regression models, particularly when extended with interaction terms, higher-order variables, or regularization techniques, can become complex and difficult to interpret. Interpretability is a significant challenge, especially in health research, where the goal is often to translate model findings into actionable insights for clinicians, policymakers, or patients. While logistic regression is generally considered more interpretable than more complex machine learning models (e.g., deep learning), challenges remain when explaining the impact of multiple predictors or interaction effects (Bakare *et al*., 2024). The log-odds interpretation of logistic regression coefficients may not always be intuitive for non-technical audiences. For example, explaining the magnitude of an odds ratio and its significance to a policymaker or healthcare provider requires clear communication strategies. To enhance interpretability, researchers have turned to visualization techniques, such as marginal effects plots or probability curves, which show how changes in predictor variables affect the outcome in a more intuitive manner. Simplified models with fewer predictors or clearer explanations of model assumptions are another strategy to make logistic regression results more accessible (Okeke *et al*., 2024). However, these efforts may come at the cost of model precision.

Finally, ethical considerations are a critical issue in logistic regression, particularly when using health data for predictive modeling (Bakare *et al*., 2024). Bias in health predictions, whether related to race, gender, or socioeconomic status, is a significant concern. Logistic regression models may inadvertently perpetuate or amplify these biases if the data used to train the model reflects historical inequities or biases in healthcare systems. For example, if a dataset is not representative of diverse populations, the model may yield inaccurate or unfair predictions for underrepresented groups. To mitigate these risks, it is crucial to ensure that health data is fairly representative and that model predictions are regularly assessed for bias. Approaches such as fairness-aware modeling and sensitivity analysis can help identify and correct biases in the model (Abass *et al*., 2024). Furthermore, the ethical use of health data is paramount. Informed consent and data privacy protections must be adhered to when using health data for research, ensuring that individuals' personal health information is handled responsibly. Transparent communication about how predictive models is used and their potential impact on health outcomes is essential for maintaining public trust and ensuring ethical integrity in health research. While logistic regression remains a valuable tool for health outcome prediction, challenges such as data issues, overfitting, interpretability, and ethical considerations must be carefully managed (Iwuanyanwu *et al*., 2024). Addressing these limitations requires thoughtful data management, appropriate model selection, and a commitment to ensuring fairness and transparency in health predictions. With continued research and methodological advancements, logistic regression will continue to play a crucial role in advancing public health and clinical decision-making.

2.4. Future Trends and Innovations in Logistic Regression for Health Outcomes

Logistic regression, a cornerstone of statistical modeling in health research, has long been instrumental in predicting health outcomes, identifying risk factors, and guiding public health interventions (Alemede *et al*., 2024). However, with advancements in technology and data analytics, logistic regression is evolving and being integrated with emerging methodologies to improve predictive accuracy, enhance interpretability, and address the growing complexity of health data. This discusses future trends and innovations in logistic regression for health outcomes, including the integration with other machine learning techniques, the role of big data, real-time predictive modeling, and improvements in model interpretability.

One of the most promising future directions for logistic regression in health outcomes is its integration with other machine learning techniques, such as random forests, support vector machines, and neural networks (Walugembe and Nakayenga, 2024) . These hybrid models aim to combine the strengths of logistic regression's simplicity and interpretability with the power of machine learning's ability to handle complex, non-linear relationships and large datasets. For instance, ensemble learning approaches, where logistic regression is used in conjunction with random forests or boosting methods, can enhance model performance by reducing overfitting and improving generalizability across diverse health datasets. Additionally, integrating neural networks with logistic regression can allow for deeper and more nuanced analysis of health data, particularly when dealing with large volumes of unstructured data, such as medical imaging or genomic information. By leveraging these hybrid models, researchers can improve predictive accuracy and achieve more robust health predictions, enhancing the ability to detect subtle patterns that may not be captured by traditional logistic regression alone.

The growing availability of big health data from sources such as electronic health records (EHRs), genomic data, and wearable health devices presents both challenges and opportunities for logistic regression models (Ezeafulukwe *et al.*, 2024; Nwaimo *et al.*, 2024). These vast datasets provide the potential for more comprehensive insights into health outcomes by incorporating a diverse array of health indicators, behaviors, and environmental factors. Logistic regression will continue to play a crucial role in analyzing these large-scale datasets, especially in predicting disease risks, identifying health determinants, and understanding trends in public health. However, handling big health data introduces challenges, including data quality issues, such as missing values, noisy data, and imbalances in the dataset (Okeke *et al.*, 2024). Additionally, scaling logistic regression to effectively analyze these large datasets requires overcoming computational limitations and ensuring data integration from multiple sources. Advances in cloud computing and distributed computing are enabling the processing of large datasets more efficiently, allowing logistic regression models to scale up and analyze real-time data streams. The future of logistic regression in big data analytics will likely involve enhanced algorithms designed to optimize performance and accuracy in the face of these challenges.

Another exciting future direction for logistic regression is its application in real-time health monitoring and decision support systems (Ekpobimi *et al.*, 2024). As wearable technologies and remote monitoring devices continue to proliferate, the ability to collect health data continuously and analyze it in real time has the potential to revolutionize personalized healthcare. Logistic regression can be used to predict health outcomes on the fly, enabling proactive interventions before health conditions worsen. For example, logistic regression models can be used in real-time to monitor heart disease risks, diabetes management, or sepsis prediction in hospitalized patients, providing healthcare providers with actionable insights as new data becomes available. Moreover, logistic regression's role in predictive analytics will continue to grow as it supports more personalized healthcare decisions (Usuemerai *et al.*, 2024). By integrating logistic regression into decision support tools, clinicians can make more informed decisions based on an individual's health data, improving patient outcomes and reducing healthcare costs. As real-time predictive modeling evolves, the potential for logistic regression to become a key component of personalized medicine where treatments are tailored to individual patients based on predictive insights will become increasingly prominent.

A significant challenge facing the future of logistic regression, and machine learning models in general, is the interpretability and transparency of complex models (Osundare and Ige, 2024). In health research and clinical settings, it is essential that predictive models are not only accurate but also understandable to healthcare professionals and policymakers. As health data and models become more complex, there is a growing need for explainable artificial intelligence (XAI) techniques that help make logistic regression models more transparent and user-friendly. Advances in model interpretability will allow healthcare providers to trust the predictions made by logistic regression models and incorporate them into clinical decision-making. The use of visualization tools, such as marginal effect plots or probability curves, will help clinicians better understand how changes in a patient's health data impact the predicted outcome (Mokogwu *et al.*, 2024; Nwaimo *et al.*, 2024). Additionally, the development of user-friendly interfaces will make it easier for healthcare professionals to interact with logistic regression models, allowing them to make more informed, evidence-based decisions.

The future of logistic regression in health outcomes holds immense potential, driven by innovations in machine learning integration, big data analytics, real-time modeling, and interpretability enhancements (Bakare *et al.*, 2024). As healthcare systems become more data-driven and personalized, logistic regression will remain a key tool for predicting health outcomes, informing clinical decisions, and guiding public health policies. However, addressing challenges such as data complexity, model transparency, and scalability will be essential to ensuring that logistic regression continues to evolve as a powerful and practical tool in the ever-changing landscape of healthcare (Ajiga *et al.*, 2024; Nwaimo *et al.*, 2024).

3. Conclusion

Logistic regression has long been a foundational statistical tool in health research, providing critical insights into disease prediction, risk factor identification, and healthcare decision-making. This has explored the core applications of logistic regression, including its use in disease prediction, public health studies, clinical decision-making, and healthcare policy formulation. The integration of logistic regression with advanced techniques like regularization, hierarchical modeling, and interaction effects has led to significant advancements in health outcome prediction. Despite its strengths, logistic regression faces challenges such as data issues, overfitting, model interpretability, and ethical concerns. The role of logistic regression in modern health research remains indispensable, particularly as it continues to support the advancement of clinical research and global health studies. By enabling researchers to model complex health data, logistic regression plays a key role in predicting outcomes for diseases like heart disease, diabetes, and cancer, as well as informing public health strategies. Additionally, it serves as a valuable tool for personalized medicine, helping clinicians predict patient-specific risks and tailor interventions accordingly. However, there is a pressing need for continued innovation in predictive modeling techniques. Future research should focus on addressing the limitations of logistic regression, such as its reliance on linearity assumptions and its vulnerability to biases in data. Exploring hybrid models that combine logistic regression with machine learning algorithms, along with the integration of big data and real-time health monitoring, could significantly enhance its predictive power. Furthermore, advancements in model transparency and interpretability will be crucial in ensuring that logistic regression models remain accessible and trustworthy for healthcare practitioners. As the field evolves, addressing these challenges will be essential for optimizing the use of logistic regression in improving global health outcomes.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Abass, L.A., Usuemerai, P.A., Ibikunle, O.E., Alemede, V., Nwankwo, E.I. and Mbata, A.O., 2024. Enhancing patient engagement through CRM systems: A pathway to improved healthcare delivery. International Medical Science Research Journal, 4(10), pp.928-960. Available at: https://doi.org/10.51594/imsrj.v4i10.1648.

[2] Achumie GO, OA Bakare, NI Okeke (2024): Implementing fair lending practices: Advanced data analytics approaches and regulatory compliance. Finance & Accounting Research Journal 2024 6 (10), 1818-1831

[3] Achumie GO, OA Bakare, NI Okeke (2024): Innovative financial and operational models for affordable housing: A review of emerging market strategies. International Journal of Applied Research in Social Sciences 2024 6 (10), 2342-2362

[4] Adekoya, O.O., Daudu, C.D., Okoli, C.E., Isong, D., Adefemi, A. and Tula, O.A., 2024. The role of environmental policies in shaping oil and gas operations: A comparative review of Africa and the USA. International Journal of Science and Research Archive, 11(1), pp.798-806.

[5] Adekoya, O.O., Isong, D., Daudu, C.D., Adefemi, A., Okoli, C.E. and Tula, O.A., 2024. Reviewing the advancements in offshore drilling technologies in the USA and their global impact. World Journal of Advanced Research and Reviews, 21(1), pp.2242-2249.

[6] Adeniran, I.A., Abhulimen, A.O., Obiki-Osafiele, A.N., Osundare, O.S., Agu, E.E. and Efunniyi, C.P., 2024. Strategic risk management in financial institutions: Ensuring robust regulatory compliance. Finance & Accounting Research Journal, 6(8), pp.1582-1596.

[7] Adeniran, I.A., Abhulimen, A.O., Obiki-Osafiele, A.N., Osundare, O.S., Agu, E.E., Efunniyi, C.P. and OneAdvanced, U.K., 2024. Data-Driven approaches to improve customer experience in banking: Techniques and outcomes. International Journal of Management & Entrepreneurship Research, 6(8), pp.2797-2818.

[8] Adewumi, A., Ewim, S.E., Sam-Bulya, N.J. and Ajani, O.B., 2024. Advancing business performance through data-driven process automation: A case study of digital transformation in the banking sector.

[9] Ajiga, D., Okeleke, P.A., Folorunsho, S.O. and Ezeigweneme, C., 2024. Methodologies for developing scalable software frameworks that support growing business needs.

[10] Ajiga, D., Okeleke, P.A., Folorunsho, S.O. and Ezeigweneme, C., 2024. Navigating ethical considerations in software development and deployment in technological giants.

[11] Akinsulire, A.A., Idemudia, C., Okwandu, A.C. and Iwuanyanwu, O., 2024. Public-Private partnership frameworks for financing affordable housing: Lessons and models. International Journal of Management & Entrepreneurship Research, 6(7), pp.2314-2331.

[12] Alemede, V., Nwankwo, E.I., Igwama, G.T., Olaboye, J.A. and Anyanwu, E.C., 2024. Pharmacists as educators: Enhancing patient understanding and access to specialty medications through community workshops. Magna Scientia Advanced Biology and Pharmacy, 13(1), pp.001-009.

[13] Alemede, V., Nwankwo, E.I., Igwama, G.T., Olaboye, J.A. and Anyanwu, E.C., 2024. Impact of 340B drug pricing program on specialty medication access: A policy analysis and future directions. Magna Scientia Advanced Biology and Pharmacy, 13(1), pp.010-018.

[14] Alemede, V., Nwankwo, E.I., Igwama, G.T., Olaboye, J.A. and Anyanwu, E.C., 2024. Designing state-level policies to support independent pharmacies in providing specialty care services in rural regions. Magna Scientia Advanced Biology and Pharmacy, 13(1), pp.019-029.

[15] Arinze, C.A., Izionworu, V.O., Isong, D., Daudu, C.D. and Adefemi, A., 2024. Integrating artificial intelligence into engineering processes for improved efficiency and safety in oil and gas operations. Open Access Research Journal of Engineering and Technology, 6(1), pp.39-51.

[16] Bakare OA, OR Aziza, NS Uzougbo, P Oduro (2024): A governance and risk management framework for project management in the oil and gas industry. Open Access Research Journal of Science and Technology 2024 12 (01), 121–13

[17] Bakare OA, OR Aziza, NS Uzougbo, P Oduro (2024): A human resources and legal risk management framework for labour disputes in the petroleum industry. Open Access Research Journal of Science and Technology, 2024, 12(01), 104–120. https://doi.org/10.53022/oarjst.2024.12.1.0118

[18] Bakare OA, OR Aziza, NS Uzougbo, P Oduro (2024): An integrated legal and business strategy framework for corporate growth in Nigerian companies. International Journal of Management & Entrepreneurship Research 6 (10), 3259. https://doi.org/10.51594/ijmer.v6i10.1614

[19] Bakare OA, OR Aziza, NS Uzougbo, P Oduro (2024): Ethical and legal project management framework for the oil and gas industry. International Journal of Applied Research in Social Sciences 2024 6 (10), 2363-2388

[20] Bakare Oluwaseun Adeola, Onoriode Reginald Aziza, Ngozi Samuel Uzougbo and Portia Oduro (2024): A legal and regulatory compliance framework for maritime operations in Nigerian oil companies. Open Access Research Journal of Science and Technology, 2024, 12(01), 092–103. https://doi.org/10.53022/oarjst.2024.12.1.0117

[21] Bakare, O.A., Aziza, O.R., Uzougbo, N.S. and Oduro, P., 2024. A legal and regulatory compliance framework for maritime operations in Nigerian oil companies.

[22] Efunniyi, C.P., Abhulimen, A.O., Obiki-Osafiele, A.N., Osundare, O.S., Agu, E.E. and Adeniran, I.A., 2024. Strengthening corporate governance and financial compliance: Enhancing accountability and transparency. Finance & Accounting Research Journal, 6(8), pp.1597-1616.

[23] Ekpobimi, H.O., 2024. Building high-performance web applications with NextJS. Computer Science & IT Research Journal, 5(8), pp.1963-1977.

[24] Ekpobimi, H.O., Kandekere, R.C. and Fasanmade, A.A., 2024. Conceptualizing scalable web architectures balancing performance, security, and usability. International Journal of Engineering Research and Development, 20(09).

[25] Ekpobimi, H.O., Kandekere, R.C. and Fasanmade, A.A., 2024. Front-end development and cybersecurity: A conceptual approach to building secure web applications. Computer Science & IT Research Journal, 5(9), pp.2154-2168.

[26] Ekpobimi, H.O., Kandekere, R.C. and Fasanmade, A.A., 2024. Software entrepreneurship in the digital age: Leveraging front-end innovations to drive business growth. International Journal of Engineering Research and Development, 20(09).

[27] Ezeafulukwe, C., Bello, B.G., Ike, C.U., Onyekwelu, S.C., Onyekwelu, N.P. and Asuzu, O.F., 2024. Inclusive internship models across industries: an analytical review. International Journal of Applied Research in Social Sciences, 6(2), pp.151-163.

[28] Ezeafulukwe, C., Onyekwelu, S.C., Onyekwelu, N.P., Ike, C.U., Bello, B.G. and Asuzu, O.F., 2024. Best practices in human resources for inclusive employment: An in-depth review. International Journal of Science and Research Archive, 11(1), pp.1286-1293.

[29] Ezeafulukwe, C., Owolabi, O.R., Asuzu, O.F., Onyekwelu, S.C., Ike, C.U. and Bello, B.G., 2024. Exploring career pathways for people with special needs in STEM and beyond. International Journal of Applied Research in Social Sciences, 6(2), pp.140-150.

[30] Ibikunle, O.E., Usuemerai, P.A., Abass, L.A., Alemede, V., Nwankwo, E.I. and Mbata, A.O., 2024. Artificial intelligence in healthcare forecasting: Enhancing market strategy with predictive analytics. International Journal of Applied Research in Social Sciences, 6(10), pp.2409–2446. Available at: https://doi.org/10.51594/ijarss.v6i10.1640.

[31] Iwuanyanwu, O., Gil-Ozoudeh, I., Okwandu, A.C. and Ike, C.S., 2024. International Journal of Applied Research in Social Sciences, 6 (8), pp. 1951-1968.

[32] Iwuanyanwu, O., Gil-Ozoudeh, I., Okwandu, A.C. and Ike, C.S., 2024. Retrofitting existing buildings for sustainability: Challenges and innovations.

[33] Iwuanyanwu, O., Gil-Ozoudeh, I., Okwandu, A.C. and Ike, C.S., 2024. The role of green building materials in sustainable architecture: Innovations, challenges, and future trends International Journal of Applied Research in Social Sciences, 6(8), pp. 1935-1950

[34] Kassem, R.G., Mbata, A.O., Usuemerai, P.A. and Ogbewele, E.G., 2022. Leveraging data analytics for optimizing pharmacy marketing strategies: Enhancing patient engagement and medication adherence. World Journal of Advanced Research and Reviews, 16(1), pp.1184-1195. Available at: https://doi.org/10.30574/wjarr.2022.16.1.1020.

[35] Kassem, R.G., Mbata, A.O., Usuemerai, P.A., Abass, L.A. and Ogbewele, E.G., 2022. Digital transformation in pharmacy marketing: integrating AI and machine learning for optimized drug promotion and distribution. World Journal of Advanced Research and Reviews, 15(2), pp.749-762. Available at: https://doi.org/10.30574/wjarr.2022.15.2.0792.

[36] Kassem, R.G., Mbata, A.O., Usuemerai, P.A., Abass, L.A. and Ogbewele, E.G., 2023. Pharmacy marketing for public health impact: Promoting preventive care and health literacy through strategic campaigns. World Journal of Advanced Research and Reviews, 18(2), pp.1406-1418. Available at: https://doi.org/10.30574/wjarr.2023.18.2.0982.

[37] Mokogwu, C., Achumie, G.O., Adeleke, A.G., Okeke, I.C. and Ewim, C.P.M., 2024. A strategic IT policy implementation model for enhancing customer satisfaction in digital markets.

[38] Mokogwu, C., Achumie, G.O., Adeleke, A.G., Okeke, I.C. and Ewim, C.P.M., 2024. A data-driven operations management model: Implementing MIS for strategic decision making in tech businesses.

[39] Mokogwu, C., Achumie, G.O., Adeleke, A.G., Okeke, I.C. and Ewim, C.P.M., 2024. A leadership and policy development model for driving operational success in tech companies.

[40] Nwaimo, C.S., Adegbola, A.E. and Adegbola, M.D., 2024. Data-driven strategies for enhancing user engagement in digital platforms. International Journal of Management & Entrepreneurship Research, 6(6), pp.1854-1868.

[41] Nwaimo, C.S., Adegbola, A.E. and Adegbola, M.D., 2024. Predictive analytics for financial inclusion: Using machine learning to improve credit access for under banked populations. Computer Science & IT Research Journal, 5(6), pp.1358-1373.

[42] Nwaimo, C.S., Adegbola, A.E. and Adegbola, M.D., 2024. Sustainable business intelligence solutions: Integrating advanced tools for long-term business growth.

[43] Nwaimo, C.S., Adegbola, A.E., Adegbola, M.D. and Adeusi, K.B., 2024. Evaluating the role of big data analytics in enhancing accuracy and efficiency in accounting: A critical review. Finance & Accounting Research Journal, 6(6), pp.877-892.

[44] Nwaimo, C.S., Adegbola, A.E., Adegbola, M.D. and Adeusi, K.B., 2024. Forecasting HR expenses: A review of predictive analytics in financial planning for HR. International Journal of Management & Entrepreneurship Research, 6(6), pp.1842-1853.

[45] Ofoegbu, K.D.O., Osundare, O.S., Ike, C.S., Fakeyede, O.G. and Ige, A.B., 2024. Proactive cyber threat mitigation: Integrating data-driven insights with user-centric security protocols.

[46] Okeke NI, OA Bakare, GO Achumie (2024): Forecasting financial stability in SMEs: A comprehensive analysis of strategic budgeting and revenue management. Open Access Research Journal of Multidisciplinary Studies, 2024 8 (1), 139-149. https://doi.org/10.53022/oarjms.2024.8.1.0055

[47] Okeke NI, OA Bakare, GO Achumie (2024): Integrating policy incentives and risk management for effective green finance in emerging markets. International Journal of Frontiers in Science and Technology Research, 7 (1), 076-088

[48] Okeke Njideka Ihuoma, Oluwaseun Adeola Bakare, and Godwin Ozoemenam Achumie (2024): Integrating policy incentives and risk management for effective green finance in emerging markets. International Journal of Frontiers in Science and Technology Research, 2024, 07(01), 076–088. https://doi.org/10.53294/ijfstr.2024.7.1.0050

[49] Osundare, O.S. and Ige, A.B., 2024. Accelerating Fintech optimization and cybersecurity: The role of segment routing and MPLS in service provider networks. Engineering Science & Technology Journal, 5(8), pp.2454-2465.

[50] Osundare, O.S. and Ige, A.B., 2024. Enhancing financial security in Fintech: Advancednetwork protocols for modern inter-bank infrastructure. Finance & Accounting Research Journal, 6(8), pp.1403-1415.

[51] Osundare, O.S. and Ige, A.B., 2024. Transforming financial data centers for Fintech: Implementing Cisco ACI in modern infrastructure. Computer Science & IT Research Journal, 5(8), pp.1806-1816.

[52] Segun-Falade, O.D., Osundare, O.S., Kedi, W.E., Okeleke, P.A., Ijomah, T.I. and Abdul-Azeez, O.Y., 2024. Assessing the transformative impact of cloud computing on software deployment and management. Computer Science & IT Research Journal, 5(8).

[53] Segun-Falade, O.D., Osundare, O.S., Kedi, W.E., Okeleke, P.A., Ijomah, T.I. and Abdul-Azeez, O.Y., 2024. Developing cross-platform software applications to enhance compatibility across devices and systems. Computer Science & IT Research Journal, 5(8).

[54] Segun-Falade, O.D., Osundare, O.S., Kedi, W.E., Okeleke, P.A., Ijoma, T.I. and Abdul-Azeez, O.Y., 2024. Evaluating the role of cloud integration in mobile and desktop operating systems. International Journal of Management & Entrepreneurship Research, 6(8).

[55] Segun-Falade, O.D., Osundare, O.S., Kedi, W.E., Okeleke, P.A., Ijomah, T.I. and Abdul-Azeez, O.Y., 2024. Developing innovative software solutions for effective energy management systems in industry. Engineering Science & Technology Journal, 5(8).

[56] Usuemerai, P.A., Ibikunle, O.E., Abass, L.A., Alemede, V., Nwankwo, E.I. and Mbata, A.O., 2024. A conceptual framework for integrating digital transformation in healthcare marketing to boost patient engagement and compliance. World Journal of Advanced Pharmaceutical and Medical Research, 7(2), pp.26–50. Available at: https://doi.org/10.53346/wjapmr.2024.7.2.0045.

[57] Usuemerai, P.A., Ibikunle, O.E., Abass, L.A., Alemede, V., Nwankwo, E.I. and Mbata, A.O., 2024. A strategic brand development framework for expanding cardiovascular and endocrinology treatments in emerging markets. World Journal of Advanced Pharmaceutical and Medical Research, 7(2), pp.77–101. Available at: https://doi.org/10.53346/wjapmr.2024.7.2.0047

[58] Usuemerai, P.A., Ibikunle, O.E., Abass, L.A., Alemede, V., Nwankwo, E.I. and Mbata, A.O., 2024. Advanced supply chain optimization for emerging market healthcare systems. International Journal of Management & Entrepreneurship Research, 6(10), pp.3321–3356. Available at: https://doi.org/10.51594/ijmer.v6i10.1637.

[59] Usuemerai, P.A., Ibikunle, O.E., Abass, L.A., Alemede, V., Nwankwo, E.I. and Mbata, A.O., 2024. A sales force effectiveness framework for enhancing healthcare access through pharmaceutical sales and training programs. World Journal of Advanced Pharmaceutical and Medical Research, 7(2), pp.51–76. Available at: https://doi.org/10.53346/wjapmr.2024.7.2.0046.

[60] Walugembe T.A and Nakayenga, N.A. (2024). Artificial Intelligence-Driven Transformation in Special Education: Optimizing Software for Improved Learning Outcomes. Association of technology and Science , 13(8), pp.163-179