DOI: 10.54254/2753-8818/36/20240614

Recent applications and perspectives of logistic regression modelling in healthcare

Qianyou Ma

School of Mathematics Statistics and Mechanics, Beijing University Of Technology, Beijing, China, 100124

934803884@qq.com

Abstract. The logistic regression model plays an important role in modeling and prediction in dealing with binary classification problems. In medical liao, the prediction, diagnosis, and treatment of diseases have always been a research direction that has attracted much attention. This paper analyses the latest applications of logistic regression models in the healthcare field from the aspects of disease risk prediction, diagnosis, and treatment outcome assessment, comparing and analyzing the latest applications and development prospects of logistic regression models in the field of healthcare, with the literature spanning 2018-2024. The paper also identifies the problems at the present stage and searches for future solutions to optimize them. Constructing logistic regression models can be effective in predicting the likelihood of disease occurrence, however, there are challenges in dealing with feature correlation and sample imbalance, as well as difficulties in dealing with missing values and outliers, and difficulties in fitting non-linear relationships, as well as difficulties in determining causality and sample imbalance. The study suggests that future research can focus on optimizing the exhibition of the calculated relapse model, including improving the feature selection method and parameter optimization algorithm of the model, or combining other algorithms and validation methods to improve the accuracy and stability of the model, and constructing a more comprehensive and diversified model to ensure the authenticity and accuracy of the data, to improve the generalization ability and the scope of application of the model.

Keywords: Data modeling, logistic regression, literature review, medicine.

1. Introduction

In recent years, with the rapid development of medical information technology and the large-scale accumulation of medical data, the application of logistic regression models in the medical field has become increasingly diverse and in-depth.

The application of logistic regression modeling in the medical field is of great significance, and it is mainly applied to three areas, namely, disease risk prediction, disease diagnosis, and treatment effect evaluation. WangFei, ZhangPing et al. proposed MulSLR: Multilinear Meager Calculated Regression.MulSLR can be viewed as a higher-request expansion of inadequate strategic relapse. A block proximal drop technique is proposed to tackle this issue and its intermingling is demonstrated. The combination pace of the proposed calculation is additionally broken down. At last, we approve the proficiency and adequacy of MulSLR in anticipating the gamble of fostering Alzheimer’s illness and cardiovascular breakdown in patients [1]. It can help doctors make accurate diagnoses, formulate personalized treatment plans and promote the progress of medical research.

Jie Ma, Paula Dhiman, et al examined how consistent indicators are dealt with in examinations creating clinical forecast models. Clinical expectation model examinations distributed between 1 July 2020 and 30 July 2020 that created strategic relapse models for parallel results were looked at in PubMed. Strategic rules are given to teach analysts the best way to deal with consistent indicators while creating clinical expectation models [2].

Steven C Bagley, Halbert White et al. analyzed the utilization and revealing of Strategic relapse in clinical writing by thoroughly evaluating its utilization in chosen areas of clinical exploration. Utilizing Medline and bibliographic pursuits, we recognized 15 companion explored articles in English distributed somewhere in the range of 1985 and 1999 and checked the articles against every one of the 10 standards for the right use and revealing of LR models. Huge lacks were tracked down in both the utilization of LR and the detailing of results. It is suggested that creators, commentators, and editors focus closely on the rules for utilizing and announcing LR models [3].

The current literature review of logistic regression modeling in healthcare focuses on a method (e.g., how to handle continuous predictors)or a model (e.g., LR model criteria). However, no paper summarises the use of logistic regression models in several major areas of the medical field and searches for room for improvement one by one. In this paper, through a literature review, we systematically sort out the latest applications of logistic regression models in the medical field, including disease risk prediction, diagnosis, and treatment effect assessment, identify the existing problems at the present stage, and look for future solutions to optimize the solutions by understanding the latest applications of logistic regression models in these medical fields. With the continuous development of technology and the strengthening of interdisciplinary cooperation, we can promote the wider and deeper development of logistic regression modeling in the medical field in the future.

2. Research status

Disease risk prediction is one of the main application areas of logistic regression models in medicine. Recent studies have shown that logistic regression models play an important role in predicting the risk of cardiovascular diseases, predicting gestational diabetes in different populations, and predicting the factors related to osteoporosis in postmenopausal women, among other research directions. By integrating individual lifestyle, genetic information and clinical data, logistic regression models can more accurately predict the probability of a patient suffering from a certain disease.

In recent years logistic regression model in the field of disease diagnosis, logistic regression has been applied in the research direction of Alzheimer's disease early diagnosis model, digital medical diagnosis, liver cancer diagnosis, and so on. In clinical practice, the logistic regression model helps doctors to diagnose diseases early by integrating multi-source information such as clinical examination, imaging and laboratory test results.

Treatment effect evaluation is another important branch of the logistic regression model in the medical field, and in recent years, it has played an important role in research directions such as the basis of treatment effect differences in nutritional epidemiology, the role of macroalgae nanoparticles in the treatment of fatty liver, and the efficacy of intravenous ketamine in the prevention of PPD. The accumulation of data from clinical trials and observational studies has provided a wealth of information for the development of logistic regression models. By analyzing patients' therapeutic responses and adverse reactions, logistic regression models can assess the efficacy and safety of different therapeutic regimens and provide a scientific basis for clinical practice.

3. Research at home and abroad

3.1. *Disease Risk Prediction*

Logistic regression models are widely used to predict patients' disease risk. Logistic regression models are constructed by analyzing the patient's clinical profile, genetic information and lifestyle factors,

DOI: 10.54254/2753-8818/36/20240614

which can effectively predict the likelihood of disease and provide decision support to physicians.

Ambrish G, Bharathi Ganesh, et al. The strategic relapse (LR) procedure was applied to characterize coronary illness in the UCI dataset. To work on the presentation of the model, the information is preprocessed, the dataset is cleaned, missing qualities are found and element determination is finished by associating every one of the highlights with the objective qualities. Exceptionally emphatically connected highlights are chosen. Then the dataset is partitioned into a preparation set for classification [4].

Managed AI prescient examination of routine antenatal consideration information from a huge well-being administration network for the period January 2016 to June 2021 was performed by Yitayeh Belsti et al. New models utilizing different indicator factors thought about factual standards and included more powerful persistent and subordinate factors. Execution measurements, including adjustment and segregation measurements, were assessed. A choice bend investigation was performed to fabricate a GDM risk expectation model by contrasting numerous AI calculations to decide the best model for anticipating GDM [5].

Yanqian Wu et al. fostered a prescient model for factors related to osteoporosis in postmenopausal ladies in the US and investigated the effect of these variables. Utilizing various relapse investigations, we discovered that age, stationary time, prednisone or cortisone use, joint pain, bone misfortune around the teeth, and rest troubles were risk factors for postmenopausal osteoporosis. Conversely, level, weight file, and progress in years finally feminine cycle were recognized as defensive factors [6].

Lucas C. Godoy MD et al incorporated all grown-up patients who went through pressure testing before going through elective coronary angiography for stable is chasmic coronary illness between April 2010 and Walk 2019 in Ontario, Canada. Applicant indicators included consolidated socioeconomics, comorbidities, research center tests, and heart stress test information. Prescient models were built utilizing customary models (strategic relapse)and AI calculations (supported trees). The plausibility of foreseeing serious left principal stenosis utilizing an extensive variety of clinical, and research facilities and painless test information was assessed [7].

Kjersti Meviket et al. surveyed the prescient legitimacy of another model that consolidates huge indicators in light of a calculated relapse model and three cutting-edge risk files to recognize high-risk patients and distinguished patients in danger for careful site contaminations by assessing the prescient legitimacy of the strategic relapse model with the gamble indices [8].

Logistic regression models are widely used to predict patients' risk of disease. By analyzing the patient's clinical data, genetic information and lifestyle, and other factors, constructing logistic regression models can effectively predict the likelihood of disease and provide decision support for doctors.

However, logistic regression models have limited ability to handle highly correlated features. When there is a strong correlation between features for disease risk prediction, logistic regression may suffer from covariance problems, leading to a decrease in the stability and generalization of the model. Moreover, in disease risk prediction, there are usually far more normal samples than risk samples, leading to sample imbalance. The logistic regression model has a limited ability to handle sample imbalance situations, which may lead to poorer model prediction for risky samples.

3.2. Disease Diagnosis

The logistic regression model is also widely used in medical imaging diagnosis. By analyzing medical imaging data, such as MRI, CT, etc., and combining it with clinical features, a logistic regression model can help doctors accurately diagnose various diseases, such as tumors, cardiovascular diseases, etc. Xiao Ruyi et al. proposed a scanty strategic relapse strategy given summed up a versatile organization for an early finding of Promotion. The summed-up flexibility network comprises Lp regularization and L2 regularisation. LP regularization creates meager solutions.L2 regularization guarantees that the applicable cerebrum districts are in the arrangement. Exploratory outcomes showed that the new strategy caught different cerebrum areas related to Promotion change and fundamentally

DOI: 10.54254/2753-8818/36/20240614

further developed Promotion order contrasted with past techniques [9].

Xinchun Cui et al. proposed a versatile Tether calculated relapse model (PSO-ALLR)given molecule swarm streamlining. The calculation is partitioned into two phases. The principal stage utilizes a molecule swarm streamlining calculation for the worldwide hunt to eliminate excess highlights and diminish the calculation time in the later stage; the subsequent stage involves versatile Rope as a neighborhood search to choose the most important elements for Promotion grouping [10].

Yousheng Zhou et al. proposed security safeguarding calculated relapse-based web-based sickness determination (LR-DDH)calculation, which safeguards the protection of clinical information through homomorphic verification encryption. Hypothetical examination and trial results show that the LR-DDH plot proposed in this paper accomplishes productive calculation and secure correspondence [11].

Juntao Li et al. fostered a versatile calculated relapse by coordinating quality transformation and RNA-seq (ALRIGMR)to resolve the issue that well-known AI-based symptomatic strategies generally overlook qualities that are not essentially differentially communicated in RNA-seq and neglect to portray covering populace impacts set off by a couple of qualities engaged with numerous natural pathways. Another information coordination system is proposed to feature qualities with high change rates and irrelevant differential articulation. Another rule for assessing quality importance in light of differential articulation and change data is proposed [12].

Thomas E. Cowling, David An et al. looked at the exhibition of calculated relapse and expanded trees in foreseeing mortality in patients with an enormous number of demonstrative codes in electronic clinical records. One-year mortality was anticipated in light of the patient’s age, sex, and financial status and whether 202 to 257 Worldwide Grouping of Illnesses, 10th Modification (ICD-10) codes (parallel indicators) had been kept in the earlier year. Strategic relapse and expanded tree demonstrating various symptomatic codes relatively anticipated patient mortality in an enormous electronic clinical record dataset [13].

In medical imaging diagnosis, a logistic regression model is also widely used. By analyzing medical image data, such as MRI, CT, etc., combined with clinical features, the logistic regression model can help doctors accurately diagnose various diseases, such as tumors, cardiovascular diseases, etc.

Be that as it may, Calculated relapse models expect a straight connection between the free and subordinate factors and don’t catch non-direct connections well. Practically speaking, illness analysis is often affected by complex non-direct factors, which might prompt unfortunate fitting of the strategic relapse model. In addition, calculated relapse models are delicate to missing qualities and exceptions, and information preprocessing is expected to adapt to these circumstances. For missing qualities, normal medicines incorporate erasing tests containing missing qualities or utilizing insertion; for exceptions, thought should be given to whether to erase or change these qualities.

3.3. Evaluation of Treatment Effectiveness

Logistic regression models also have important applications in predicting patient response to drugs. Logistic regression models are constructed to predict a patient’s response to a specific drug by analyzing the patient’s genotype, drug metabolism and other factors to individualize treatment.

Gomila, Robin utilized the econometric hypothesis and laid out measurements to show that straight relapse is much of the time the best technique for assessing the causal impact of a treatment on a double result. Straight relapse coefficients can be deciphered straightforwardly in probabilities, and direct relapse is more secure when connection terms or fixed impacts are incorporated. The Neyman-Rubin causal model is evaluated and used to systematically exhibit that direct relapse produces impartial appraisals of treatment results on double results. Clinicians are encouraged to utilize straight relapse to appraise treatment consequences for parallel results [14].

Shu Teng, Zheng Nan et al. gathered information on the restorative worth of macroalgae nanoparticles for greasy liver through trial displaying. This study was meant to assess the defensive impact of kelp, Pardina pavonia (PP), against carbon tetrachloride (CCl_4)- prompted liver fibrosis. In this review, two models, calculated relapse (LR)and backing vector machine (SVM), were proposed to foresee the probability of liver illness advancement. The model was approved for research center tests anticipating sickness (progress in years; direct bilirubin (DB), complete protein (TP)and egg whites (ALB). The execution time and arrangement exactness of these classifiers were analyzed [15].

Judith J. M. Rijnhart, Jos W. R. Twisk et al. thought about the presentation of impact gauges created by the three techniques utilizing reproduction studies and two genuine information (n = 360)from an observational companion study, exhibiting unstandardized and relative circuitous impacts and extents in light of the potential result systems of multivariate relapse, primary condition displaying, and intervention models with dichotomous results. normalized gauges. Relative execution of normalized and standardized gauges [16].

Shen Pei and Liu Xiaohan et al. investigated the effect of bone remodeling after arthroscopic treatment of intervertebral disc displacement by retrospective analysis. Methods:1850 patients with anterior disc displacement of the temporomandibular joint admitted from December 2010 to December 2016 were selected. Magnetic resonance imaging (MRI)was performed preoperatively and postoperatively. The occurrence of bone changes after arthroscopic disc repositioning treatment was analyzed and compared. A logistic regression algorithm was used to measure the relevant factors affecting new bone formation after surgery [17].

The logistic regression model also has important applications in predicting patients' responses to drugs. By dissecting the patient's genotype, drug digestion and different elements, the development of a calculated relapse model can foresee the patient's reaction to explicit medications, to accomplish individualized treatment.

Logistic regression models are essentially associative models that have difficulty capturing causal relationships. Treatment outcome evaluations often need to consider other underlying factors and interventions, and logistic regression models cannot directly address these causal relationships. Moreover, there are often positive and negative sample imbalances in treatment outcome evaluations, and logistic regression models are limited in their ability to handle sample imbalances, which may result in predictions that are biased toward categories with higher proportions.

4. Conclusion

This paper mainly summarises the application of logistic regression models in three areas of the medical field, namely, disease risk prediction, disease diagnosis and treatment effect evaluation, using the method of literature review. By understanding the latest applications of logistic regression models in these medical fields, and identifying the existing problems and solutions for optimisation and improvement through the literature review, but in terms of the cited literature, the number of literature for the three fields as examples is relatively small, and it fails to completely cover all the cutting-edge applications.

Future research could focus on optimising the performance of the logistic regression model, including improving the feature selection methods and parameter optimisation algorithms of the model or combining other algorithms and validation methods to improve the accuracy and stability of the model. With the continuous accumulation of medical data, future research could try to merge datasets from different sources to construct a more comprehensive and diverse model to ensure the authenticity and accuracy of the data, thus improving the generalisation ability and applicability of the model. Future research could also explore the option of combining logistic regression models with other emerging technologies to further improve the performance and application range of the models by incorporating cutting-edge knowledge from other disciplines.

References

[1] Jie Ma, Paula Dhiman et al., (2023). Poor handling of continuous predictors in clinical prediction models using logistic regression: a systematic review,Journal of Clinical Epidemiology,Volume 161,Pages 140-151.
[2] Steven C Bagley, Halbert White et al,. (2001). Logistic regression in the medical literature::

DOI: 10.54254/2753-8818/36/20240614

Standards for use and reporting, with particular attention to one medical domain,Journal of Clinical Epidemiology,Volume 54, Issue 10,Pages 979-985.

[3] Wang, Fei and Zhang, Ping et al,. (2014). Clinical risk prediction using multilinear sparse logistic regression,Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,KDD '14,145–154.

[4] Ambrish G, Bharathi Ganesh et al., (2022). Logistic regression technique for prediction of cardiovascular disease,Global Transitions Proceedings,Volume3, Issue1, Pages127-130.

[5] Yitayeh Belsti, Lisa Moran et al., (2023). Comparison of machine learning and conventional logistic regression-based prediction models for gestational diabetes in an ethnically diverse population; the Monash GDM Machine learning model,International Journal of Medical Informatics,Volume 179,105228.

[6] Yanqian Wu, Jianqian Chao et al., (2023). Construction of predictive model for osteoporosis-related factors among postmenopausal women on the basis of logistic regression and Bayesian network,Preventive Medicine Reports,Volume 35,102378.

[7] Lucas C. Godoy, Michael E. Farkouh, et al., (2023). Predicting left main stenosis in stable ischemic heart disease using logistic regression and boosted trees,American Heart Journal,Volume 256,Pages 117-127.

[8] Kjersti Mevik, Ashenafi Zebene Woldaregay et al., (2024). Exploring surgical infection prediction: A comparative study of established risk indexes and a novel model,International Journal of Medical Informatics,Volume 184,105370.

[9] Ruyi Xiao, Xinchun Cui et al., (2021). Early diagnosis model of Alzheimer's disease based on sparse logistic regression with the generalized elastic net,Biomedical Signal Processing and Control, Volume66,102362.

[10] Xinchun Cui, Ruyi Xiao et al., (2021). Adaptive LASSO logistic regression based on particle swarm optimization for Alzheimer's disease early diagnosis,Chemometrics and Intelligent Laboratory Systems,Volume 215,104316.

[11] Yousheng Zhou, Liyuan Song et al., (2023). A privacy-preserving logistic regression-based diagnosis scheme for digital healthcare,Future Generation Computer Systems,Volume 144,Pages 63-73.

[12] Juntao Li, Fuzhen Cao et al,. (2024). ALRIGMR: Adaptive logistic regression via integrating gene mutation and RNA-seq for liver cancer diagnosis,Biomedical Signal Processing and Control,Volume 91,106025.

[13] Thomas E. Cowling, David A. Cromwell et al,. (2021). Logistic regression and machine learning predicted patient mortality from large sets of diagnosis codes comparably,Journal of Clinical Epidemiology, Volume 133, Pages 43-52.

[14] Gomila, R. (2021). Logistic or linear? Estimating causal effects of experimental treatments on binary outcomes using regression analysis,Journal of Experimental Psychology: General, 150(4), 700–709.

[15] Shu Teng, Nan Zheng et al., (2023). Effect of nanoparticle macroalgae in the treatment of fatty liver disease using logistic regression, and support vector machine,Environmental Research, Volume 224,115426.

[16] Julie M. Petersen, Ashley I. Naimi et al., (2023). Does heterogeneity underlie differences in treatment effects estimated from SuperLearner versus logistic regression? An application in nutritional epidemiology,Annals of Epidemiology,Volume 83,Pages 30-34.

[17] Shen Pei, LIU Xiaohan et al., (2018). Analysis of the effect of arthroscopy on bone remodeling after disk displacement [A]. Temporomandibular Joint and Dentistry Committee of Chinese Stomatological Association. Compilation of papers of the 15th National Symposium on Temporomandibular Joint and Dentistry of Chinese Stomatological Association [C]. Shanghai Key Laboratory of Stomatology, Department of Oral Surgery, The Ninth People's Hospital of Stomatology, Shanghai Jiaotong University School of Medicine; : Chinese Academy of Stomatology.