MATHEMATICAL MODELING AND COMPUTING, Vol. 12, No. 1, pp. 132–143 (2025)

Enhancing logistic regression model through AHP-initialized weight optimization using regularization and gradient descent adaptation: A comparative study

Kamal A. A.[1,2], Mansor M. A.[1], Kasihmuddin M. S. M.[3]

[1] *School of Distance Education, University Sains Malaysia, 11800 USM, Malaysia*
[2] *Centre for Pre-University Studies, University Malaysia Sarawak, 94300 Kota Samarahan, Malaysia*
[3] *School of Mathematical Sciences, University Sains Malaysia, 11800 USM, Malaysia*

(Received 15 October 2024; Revised 7 February 2025; Accepted 15 February 2025)

This study explores an approach to improving the performance of logistic regression model (LR) integrated with Analytic Hierarchy Process (AHP) for weight initialization model with regularization and adaptation of gradient descent (GD). Traditional LR model relies on random weight initialization leading to suboptimal performances. By employing AHP, a hybrid model that deployed priority vector as initial weights is obtained, reflecting the relative importance of input features. Previous works reported subpar performances of AHP-LR hybrid model due to the lack of optimizing for the initialized weights. In this study, the weights are proposed to be optimized with L1 and L2 regularization approach, penalizing deviations from the AHP-initialized weights through modified log-likelihood function with modified GD optimization. This comparative analysis involves four models: LR with L2 regularization, AHP weights as LR weights, and AHP-weights optimized with L1 and L2 regularization. A prediction experiment is conducted using synthetic dataset to assess the models' performance in terms of accuracy, recall, precision, F1-score, and ROC-AUC. The results indicate that optimizing weights with L1 or L2 regularization significantly enhances model performance, compared to direct application of AHP weights without optimization yields near-random guesses. Additionally, incorporating true expert-derived weights, evaluating their impact on model performance and experimenting with authentic dataset and different weight derivation methods would offer valuable insights.

Keywords: *logistic regression; analytical hierarchy process; hybrid model; regularization; modified gradient descent.*

2010 MSC: 62J12 **DOI:** 10.23939/mmc2025.01.132

1. Introduction

In the advancement of statistical modeling and data analysis, integrating machine learning (ML) with multi-criteria decision-making (MCDM) techniques provides potential for research on enhancing the hybrid model. This study explores this hybridization by utilizing the analytical hierarchy process (AHP) for logistic regression (LR) weight adjustment. This integration aims to address the limitations of conventional weight assignment methods [1] to enhance the ML model's predictive power and applicability. This methodological advancement improves accuracy and facilitates more precise data interpretation in various domains where LR is pivotal [2–5]. As example, an improved LR models could lead to better investigations of factors contributing to disease symptoms [6] and predict loss given default in finance [7]. Fields such as esports would also benefit significantly from the application of enhanced LR models [8], particularly with AHP's strength in domain-specific decision-making [9]. Broadly, this study contributes to ongoing efforts to improve statistical methodologies through interdisciplinary approaches.

This work was supported by both USM and UNIMAS. Their invaluable support offers a stimulating academic environment that significantly contributed to the completion of this research.

LR stands out as a fundamental ML model widely applied across diverse fields [10], ranging from medical research [11,12] to life [13] and social sciences [14]. The LR model is renowned for its ability to model binary outcomes and has become crucial in interpreting complex data structures. Despite its vast applications, traditional LR faces inherent challenges, particularly in the weighting and prioritization of predictor variables [15–17]. AHP is a structured technique for organizing and analysing complex decisions [18]. AHP, a tool in operations research, provides a robust framework for dealing with complex situations where multiple criteria must be weighed [19–22]. AHP presents a novel approach to refining ML algorithms [23].

Previous works [24–26] have highlighted the need for a flexible, non-rigid model due to the limitations of data-driven AHP initial weights, which lack the interpretations that human experts can provide. Despite the advancements, a significant research gap remains in effectively integrating AHP-initialized weights with LR. This gap can be addressed by proposing an optimization on the AHP derived weights with regularization and gradient descent adaptation. The log-likelihood function and the gradient descent (GD) training phase must be modified by adding a penalty term in the form of a regularization parameter. The significance of this study lies in its potential to bridge this gap, offering a more accurate and interpretable LR model. The research objectives are threefold: (i) to formulate an L1 and L2 regularization approach to the hybrid model, (ii) to propose a modified gradient descent algorithm for the weights optimization, and (iii) to compare the effectiveness of the proposed method.

2. Theoretical background

2.1. Logistic regression

The interpretability of LR allows for a transparent understanding of how each feature influences predictions, a crucial aspect for the AHP integration [27]. LR models the probability of an event taking place by having the log-odds for the event be a linear combination of one or more independent variables. LR is a linear combination of input features that needs to be transformed by logistic function [28]. The simplest LR has one input feature x_i, with its weight w_i and one dependent outcome y_i, where $i = 1$, as shown in (1). The b is y-intercept, defined as the bias term,

$$y_1 = w_1 x_1 + b. \tag{1}$$

Equation (1) shows the fundamental linear regression model which requires logistic function characteristics to make prediction for complex, authentic problems [29]. Hence, the sigmoid function, with arbitrary parameter z, shown in (2), is applied to transform (1) to better fit the dataset by transforming the real values of y_i, mapping them to the interval [30–32],

$$\sigma(z_i) = \frac{1}{1 + e^{-z_i}}. \tag{2}$$

The transformed function is the example of LR model for one input feature, shown in (3),

$$y_1 = \sigma(w_1 x_1 + b). \tag{3}$$

Authentic problems commonly comprise of multiple input features, x_n, as in (4) [33–37]. Multiple input features have equivalent numbers of feature weights, w_n,

$$Y = w_n x_n + w_{n-1} x_{n-1} + \ldots + w_1 x_1 + b. \tag{4}$$

Equation (4) can be rewritten in a simple matrix form, as shown in (5),

$$Y = W^T X + b, \quad \text{where} \quad W^T X = \begin{bmatrix} w_n x_n \\ w_{n-1} x_{n-1} \\ \vdots \\ w_1 \end{bmatrix}. \tag{5}$$

By applying sigmoid function into (5), the LR model in (6) can be used to make prediction for any number of datasets based on the probability, $p(x_i)$,

$$p(x_i) = \sigma(z_i) = \sigma\left(W^T x_i + b\right) = \frac{1}{1 + e^{-(W^T x_i + b)}}. \tag{6}$$

From (6), the parameter to be optimized is the matrix W. Weight optimization for the LR model is done by finding the values for weights that minimizes the error based on training data using cost function. The most common cost function is usually the log-loss function [38], shown in (7), where m is the number training samples, and y_i is the true label of the ith sample. This function is chosen to ensure the study enable the observation of the proposed approach's effects towards the performance of the model, rather than the sophistry of the function used,

$$J(W) = -\frac{1}{m}\sum_{i=1}^{m}\left[y_i \log\big(p(x_i)\big) + (1-y_i)\log\big(1-p(x_i)\big)\right]. \tag{7}$$

2.2. Gradient descent

GD is a common optimization algorithm used to find the minimum of a function by iteratively adjusting the weights of the model [39]. The algorithm works by computing the gradient of the cost function with respect to its weights and bias term, which gives the direction of steepest ascent. The weights commonly initialized as random values or zeroes [40] are then updated to move towards a minimum of the cost function [41].

Algorithm 1 Gradient descent

1: Initialize the weights, W.
2: Set learning rate α.
3: **repeat**
4: Compute the gradient of cost function with respect to weights and bias, as shown in (8) and (9) respectively

$$\frac{\partial J}{\partial w_i} = -\frac{1}{m}\sum_{i=1}^{m}\big[y_i - p(x_i)\big]\cdot x_i, \tag{8}$$

$$\frac{\partial J}{\partial b} = -\frac{1}{m}\sum_{i=1}^{m}\big[y_i - p(x_i)\big]. \tag{9}$$

5: Update parameters: W and b, as shown in (10), (11), and (12)

$$w_{i,new} = w_i - \alpha\cdot\frac{\partial J}{\partial w_i}, \tag{10}$$

$$b_{new} = b - \alpha\cdot\frac{\partial J}{\partial b}, \tag{11}$$

$$w_i = w_{i,new}, \quad b = b_{new}, \tag{12}$$

6: **until** convergence criterion is met.

2.3. Analytical hierarchy process

The problem of interest is defined as criteria that will be used to evaluate the decision in a hierarchical structure called a hierarchy tree diagram as in Figure 1. The top of the tree is the goal of decision-making, followed by criteria and sub-criteria, then the alternatives that must be considered through pairwise comparison.

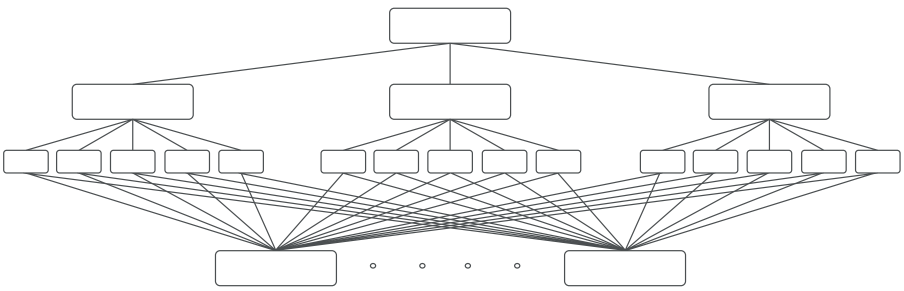

Fig. 1. The hierarchy structure of the decision-making problem, the top is the goal.

Pairwise comparisons are made using the scale as shown in Table 1. Then, we calculate the weights for each element in the pairwise comparison matrix that represent their relative importance.

Table 1. The fundamental scale of AHP [42].

Values assigned	Meaning of relative importance
1	Equal importance
3	Moderate importance
5	Strong importance
7	Very strong importance
9	Absolute importance
$2, 4, 6, 8$	For compromise between above-specified value

A is a reciprocal, square matrix with each element is represented as $a_{i,j}$ which indicates how important criterion in any column i compared to criterion in any row j, as shown in (13),

$$A = \begin{bmatrix} 1 & a_{1,2} & \cdots & \cdots & a_{1,j} \\ \frac{1}{a_{1,2}} & 1 & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & a_{i-1,j-1} \\ \frac{1}{a_{1,j}} & \cdots & \cdots & \frac{1}{a_{i-1,j-1}} & 1 \end{bmatrix}. \tag{13}$$

We would be able to compute priority vector $\mathbf{w}$, the normalized Eigen vector of A. The priority vector shows relative weights among the compared criteria and alternatives. Then, the measure of consistency called the consistency index, CI is given in (14),

$$CI = \frac{\lambda_A - n}{n-1}. \tag{14}$$

To use the CI, it is compared to the random consistency index, RI by using the formula shown in (15). This formula is called consistency ratio,

$$CR = \frac{CI}{RI}. \tag{15}$$

The values of RI are as shown in Table 2 based on the work by Prof. Saaty. If given that, the consistency of comparison made is acceptable. Otherwise, there is a need to revise the expert's judgment on the pairwise comparison.

Table 2. RI found to measure CI [42].

n	1	2	3	4	5	6	7	8	9	10
RI	0.00	0.00	0.58	0.9	1.12	1.24	1.32	1.41	1.45	1.49

2.4. Hybrid model

The AHP-LR hybrid model consists of employing AHP to assign weights to the input features. The LR model then performed prediction task with these AHP-derived weights [43]. Based on the findings, optimization is not performed on the obtained weights [44], as represented in (16),

$$W = \mathbf{w}. \tag{16}$$

It is also highlighted that observations were recorded and weighted using AHP without the necessity of performing pairwise comparisons with human experts' judgments [45]. Despite studies might not explicitly mentioned, typically some limitations that may be faced in the hybrid model include dependence on the accuracy of input data [46], the rigid nature of the hybrid models [47], which might not account for the generalization and flexibility of the model, and the potential for over-fitting or under-fitting.

3. Method

3.1. Model formulation

3.1.1. Regularization implementation

Two types of regularization, L1 and L2 can be implemented as in (17) and (18) respectively by introducing a penalty term, λ_1 and λ_2 that controls the penalization of the weights' magnitude the further they deviate from the AHP derived weights, with w_j representing the weights for the jth input feature [48],

$$\text{L1 regularization term} = \lambda_{l1} \sum_{j=1}^{n} |w_j|, \tag{17}$$

$$\text{L2 regularization term} = \lambda_{l2} \sum_{j=1}^{n} w_j^2. \tag{18}$$

3.1.2. Modified objective function

The objective function is the regularized cost function, denoted by the log-loss minus the penalty term. As optimization algorithms usually minimize functions, so the cost function is often considered negative of log-loss added with the regularization term. Here, we denote the cost function for L1 regularization as J_1, and L2 regularization as J_2, as shown in (19) and (20), respectively,

$$J_1(W) = -\frac{1}{m} \sum_{i=1}^{m} \Big[y_i \log\big(p(x_i)\big) + (1-y_i) \log\big(1-p(x_i)\big) \Big] + \lambda_{l1} \sum_{j=1}^{n} |w_j|, \tag{19}$$

$$J_2(W) = -\frac{1}{m} \sum_{i=1}^{m} \Big[y_i \log\big(p(x_i)\big) + (1-y_i) \log\big(1-p(x_i)\big) \Big] + \lambda_{l2} \sum_{j=1}^{n} w_j^2 \tag{20}$$

3.1.3. Modified gradient function

Due to the modification in objective function, the gradient function in the GD training phase needs to be adjusted. Then, the gradient of the cost function with respect to the weights is given by (21) and (22) for L1 and L2 regularization respectively. This change affects the update phase in the GD iterative step as well,

$$\frac{\partial J_1}{\partial w_j} = -\frac{1}{m} \sum_{i=1}^{m} \big[y_i - p(x_i) \big] \cdot x_i + \lambda_{l1} \cdot g(w_j), \tag{21}$$

where

$$g(w_j) = \begin{cases} 1, & w_j > 0, \\ -1, & w_j < 0, \\ 0, & w_j = 0, \end{cases}$$

$$\frac{\partial J_2}{\partial w_j} = -\frac{1}{m} \sum_{i=1}^{m} \big[y_i - p(x_i) \big] \cdot x_i + \frac{\lambda_{l2}}{2} w_j. \tag{22}$$

3.2. Model assessment

3.2.1. Models implementation

This study involved assessment of four distinct models to assess the impact of AHP-derived weights and regularization techniques on LR performance. The models are defined as follows:

— **Model I**: LR model with L2 regularization applied directly to the predictor variables, serving as a baseline model to compare the effects of AHP-derived weights and regularization.

— **Model II**: LR model where AHP-derived weights were used directly as the regression coefficients without any further optimization or regularization, to assess the efficacy of AHP weights in their raw form.

- **Model III**: LR model in which the AHP-derived weights were optimized using L1 regularization, aiming to enhance model performance by imposing sparsity on the coefficients.
- **Model IV**: LR model where the AHP-derived weights were optimized with L2 regularization, like Model III, but with the focus on minimizing over-fitting by penalizing large coefficients.

3.2.2. Confusion matrix

The confusion matrix is a table that is often used to describe the performance of a classification model on a set of test data for which the true values are known [49]. It is a way to visualize the accuracy of a classifier by comparing the predicted values with the true values. In binary classification, the confusion matrix is a 2×2 table that contains four entries: true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). These entries represent the number of instances in each of the following categories:

- **TP**: instances that are positive and were correctly classified as positive by the model.
- **FP**: instances that are negative but were incorrectly classified as positive by the model.
- **TN**: instances that are negative and were correctly classified as negative by the model.
- **FN**: instances that are positive but were incorrectly classified as negative by the model.

The confusion matrix will be used to compute various performance metrics, such as accuracy, precision, recall, and F1 score, which provide different perspectives on the performance of the classifier.

3.2.3. Accuracy

Accuracy in machine learning is the percentage of correctly classified instances out of the total instances evaluated. The accuracy of the model can be calculated using the formula in (23),

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \tag{23}$$

3.2.4. Area under the receiver operating characteristics curve

The AUC-ROC curve is a widely used metric for evaluating the performance of a binary classification model that provides both the visual and quantitative measurements to represent the performance of model. The ROC curve is a graphical representation of a classification model's diagnostic ability that plots the True Positive Rate (TPR) against the False Positive Rate (FPR) at various threshold settings. TPR, also known as sensitivity or recall, measures the proportion of actual positives that are correctly identified by the model. FPR measures the proportion of actual negatives that are incorrectly identified as positives. The AUC measures the 2-dimensional area under the ROC curve to provide a cumulative measure of the model's performance across all possible classification thresholds. The interpretation of AUC values is given in Table 3.

Table 3. Interpretation of AUC values [50].

AUC value	Interpretation
0.90 – 1.00	Excellent
0.80 – 0.90	Good
0.70 – 0.80	Fair
0.60 – 0.70	Poor
0.50 – 0.60	Fail

4. Results and discussion

4.1. Dataset

The dataset used for the experimental setup in this study is called *Santander Customer Transaction Prediction*. The dataset is available for online use via Kaggle repository (link to the data https://www.kaggle.com/competitions/santander-customer-transaction-prediction/overview). The dataset in the train CSV file consists of 200 000 observations with 200 features. In this study, we used only 100 000 observations that are splitted into a common ratio of 70% train and 30% test dataset [51].

4.2. Analytical hierarchy process derived weights

As the dataset consists of records for customers' transaction as target (dependent outcome) based on arbitrary, unidentified features (independent variables), they are not linked to a specific domain.

Table 4. Several AHP simulated weights values.

Features	Simulated weights
1	0.006441
2	0.006769
3	0.007386
...	...
200	0.004516

Therefore, it is infeasible to generate true expert AHP-derived weights. This is a limitation for real-world dataset, but it is suitable to obtain synthetic AHP weights that can be served as a proof of concept. Thus, we simulated expert input approach to create a more structured and data-informed AHP weighting system which enhance the credibility of the weights even without true domain expertise. First, the key features were identified by calculating statistical measures which are correlation and variance for each feature relative to the target variable. This simulates how an expert might prioritize features based on their impact. Next, these measures were used to generate pairwise comparison scores, to mimic the typical AHP expert judgment process, where features with higher correlation to the target are given higher comparison scores. A pairwise comparison matrix was constructed using ratios of correlations and variances, and the AHP weights and CR were calculated. Table 4 shows AHP weights values for several features with $CR < 0.10$.

4.3. Models' performances

Table 5. Train and test AUC scores for the models.

Metrics	Model I	Model II	Model III	Model IV
Train AUC	0.8617	0.4848	0.8632	0.8632
Test AUC	0.8596	0.4974	0.8613	0.8613

Four models are deployed with the dataset and the AUC scores are measured. The scores are given in Table 5. High scores are good indicator of the models' performances, but the difference between AUC score from training and testing dataset indicates that the model may be under-fitting or over-fitting [52–54]. Model I train and test AUC values are close, indicating the model generalizes well and does not exhibit apparent over-fitting as L2 regularization is effective in controlling dataset complexity [55]. Model III and IV display better scores, suggesting good balance between complexity and regularization strength. Using AHP weight without optimization, however, yield scores close to 0.5, indicating near random guessing and predictions.

Table 6. Performance metrics across all model.

Metrics	Model I	Model II	Model III	Model IV
Accuracy	0.914500	0.498733	0.915700	0.915733
Precision	0.900815	0.818973	0.902343	0.902398
Recall	0.914500	0.498733	0.915700	0.915733
F1-score	0.902757	0.594310	0.897595	0.897650

Table 6 displays the assessment of the models' accuracy, precision, recall, and F1-score. Figure 2 shows the combined ROC curves from all models deployed. Model I shows strong overall performance, achieving a high precision of 0.9008, recall of 0.9145, and F1-score of 0.9028. This model maintains high level values across all metrics, suggesting robust ability to identify positive instances and minimizing false positives.

Model II has noticeably lower performance metrics and falls significantly short as compared to other models. These results suggest that without further optimization or regularization, the AHP-initialized weights do not align sufficiently with the dataset's underlying patterns, leading to near-random performance levels.

Both Model III and IV show improvements over Model II, showing that optimization plays a vital role in adapting AHP weights effectively. Models III and IV achieve higher precision compared to Model I and II. Model IV closely matches the excellent results of Model III, with slightly lower precision but a marginally higher recall and accuracy. Figure 3 shows the confusion matrix of the test dataset from Model III and IV. Both models display strong, balanced performance, demonstrating the effectiveness of regularization techniques in optimizing AHP-derived weights. AHP-initialized weights in LR model alone are not sufficient to be deployed in predictive modeling, and regularization may significantly enhance the hybrid model efficacy. Model I serves as a strong baseline, but the optimized AHP models demonstrate the potential of incorporating expert-derived weights, provided they are adequately adjusted using regularization techniques which help in preventing over-fitting and achieve better generalization on test data.

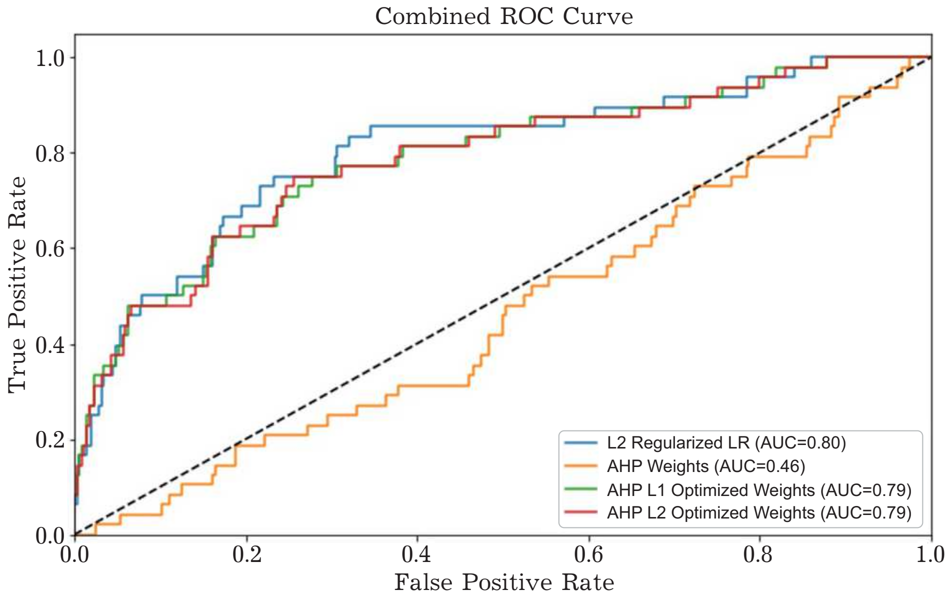

Fig. 2. The ROC curves of all four models.

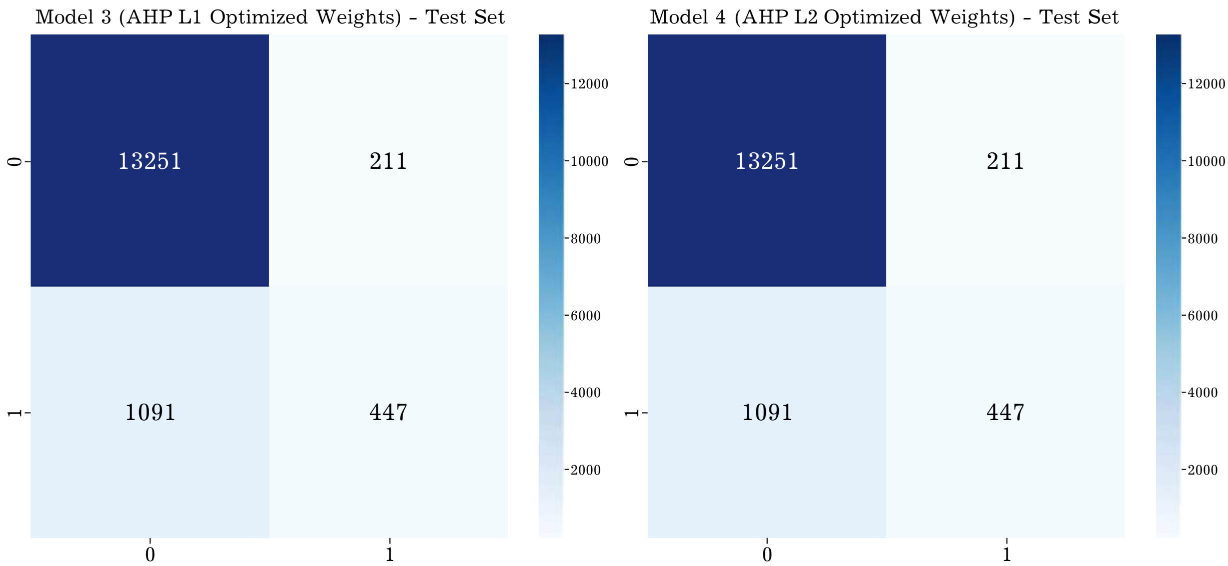

Fig. 3. The confusion matrix for model IV from train and test dataset.

5. Conclusions

This study assessed the performance of the AHP-LR model with regularization approach. The formulation of an L1 and L2 regularization approach to the hybrid model was conducted successfully. A modified gradient descent algorithm was proposed for the purposed of optimizing AHP-initialized weights and a comparative analysis for the effectiveness of the proposed method was made tactfully. The results indicate that optimizing AHP weights with L1 or L2 regularization significantly enhances model performance and mitigates over-fitting issue, whereas direct application of AHP weights without optimization is ineffective, yielding near-random predictions.

Despite the promising findings, this study has several limitations. The analysis was conducted on large, synthetic dataset, independent of specific domain may not fully capture the complexities of authentic problems. Additionally, the model used was of LR model, potentially limiting their ability to capture complex relationships within the data. The AHP weights were derived using statistical measures of the dataset rows to simulate how experts perform judgment on features' importance, which may not represent true expert knowledge, possibly affecting the efficacy of these weights.

Future work should focus on addressing these limitations. Conducting similar studies on larger and more complex datasets would validate the findings and explore the scalability of the proposed methods. Further exploration of hybrid approaches such as Elastic Net, which combines L1 and L2 regularization could be valuable. This would refine AHP's alignment with dataset-specific feature importance, enhancing generalizability without over-fitting. Implementing more advanced ML models, such as neural networks or ensemble methods, could provide a comparison with LR performance. Incorporating true expert-derived AHP weights and evaluating their impact on model performance, experimenting with different weight derivation methods, would offer valuable insights.

[1] Kim B., Shin S. J. Principal weighted logistic regression for sufficient dimension reduction in binary classification. Journal of the Korean Statistical Society. **48** (2), 194–206 (2019).

[2] Huang F., Cao Z., Guo J., Jiang S.-H., Li S., Guo Z. Comparisons of heuristic, general statistical and machine learning models for landslide susceptibility prediction and mapping. CATENA. **191**, 104580 (2020).

[3] Bui D. T., Tsangaratos P., Nguyen V.-T., Liem N. V., Trinh P. T. Comparing the prediction performance of a Deep Learning Neural Network model with conventional machine learning models in landslide susceptibility assessment. CATENA. **188**, 104426 (2020).

[4] Pham B. T., Jaafari A., Prakash I., Bui D. T. A novel hybrid intelligent model of support vector machines and the MultiBoost ensemble for landslide susceptibility modeling. Bulletin of Engineering Geology and the Environment. **78** (4), 2865–2886 (2019).

[5] Wu Y., Ke Y., Chen Z., Liang S., Zhao H., Hong H. Application of alternating decision tree with AdaBoost and bagging ensembles for landslide susceptibility mapping. CATENA. **187**, 104396 (2020).

[6] Ruan Z., Li D., Cheng X., Jin M., Liu Y., Qiu Z., Chen X. The association between sleep duration, respiratory symptoms, asthma, and COPD in adults. Frontiers in Medicine. **10**, 1108663 (2023).

[7] Breed D. G., Verster T., Schutte W. D., Siddiqi N. Developing an Impairment Loss Given Default Model Using Weighted Logistic Regression Illustrated on a Secured Retail Bank Portfolio. Risks. **7** (4), 123 (2019).

[8] Maymin P. Z. Smart kills and worthless deaths: eSports analytics for League of Legends. Journal of Quantitative Analysis in Sports. **17** (1), 11–27 (2021).

[9] Chen Y., Annebicque D., Philippot A., Carré-Ménétrier V., Daneau T. Evaluation Methodology of Interoperability for the Industrial Domain: Standardization vs. Mediation. Processes. **11** (4), 1274 (2023).

[10] Chen C., Zhou J., Wang L., Wu X., Fang W., Tan J., Wang L., Liu A. X., Wang H., Hong C. When Homomorphic Encryption Marries Secret Sharing: Secure Large-Scale Sparse Logistic Regression and Applications in Risk Control. Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. 2652–2662 (2021).

[11] Renganathan V. Overview of artificial neural network models in the biomedical domain. Bratislava Medical Journal. **120** (7), 536–540 (2019).

[12] Boateng E. Y., Abaye D. A. A Review of the Logistic Regression Model with Emphasis on Medical Research. Journal of Data Analysis and Information Processing. **7** (4), 190–207 (2019).

[13] Sharma V., Hong T., Cecchi V., Hofmann A., Lee J. Y. Forecasting weather-related power outages using weighted logistic regression. IET Smart Grid. **6** (5), 470–479 (2023).

[14] Kang S. K., Peng L., Xiao H. Risk analysis with categorical explanatory variables. Insurance: Mathematics and Economics. **91**, 238–243 (2020).

[15] Alsmadi I., Hoon G. K. Term weighting scheme for short-text classification: Twitter corpuses. Neural Computing and Applications. **31** (8), 3819–3831 (2019).

[16] Trinh T., Luu B. T., Le T. H. T., Nguyen D. H., Van Tran T., Van Nguyen T. H., Nguyen K. Q., Nguyen L. T. A comparative analysis of weight-based machine learning methods for landslide susceptibility mapping in Ha Giang area. Big Earth Data. **7** (4), 1005–1034 (2022).

[17] Maalouf M., Homouz D., Trafalis T. B. Logistic regression in large rare events and imbalanced data: A performance comparison of prior correction and weighting methods. Computational Intelligence. **34** (1), 161–174 (2018).

[18] Liu Y., Eckert C. M., Earl C. A review of fuzzy AHP methods for decision-making with subjective judgements. Expert Systems with Applications. **161**, 113738 (2020).

[19] Darko A., Chan A. P. C., Ameyaw E. E., Owusu E. K., Pärn E., Edwards D. J. Review of application of analytic hierarchy process (AHP) in construction. International Journal of Construction Management. **19** (5), 436–452 (2019).

[20] Kusherbaeva V., Zhou N. Application of the analytic hierarchy process to plant operation optimization. 2019 International Conference on Computational Science and Computational Intelligence (CSCI). 1196–1202 (2019).

[21] Stofkova J., Krejnus M., Stofkova K. R., Malega P., Binasova V. Use of the Analytic Hierarchy Process and Selected Methods in the Managerial Decision-Making Process in the Context of Sustainable Development. Sustainability. **14** (18), 11546 (2022).

[22] Repetski E., Sarkani S., Mazzuchi T. Applying The Analytic Hierarchy Process (AHP) To Expert Documents. International Journal of the Analytic Hierarchy Process. **14** (1), 1–14 (2022).

[23] He H., Hu D., Sun Q., Zhu L., Liu Y. A Landslide Susceptibility Assessment Method Based on GIS Technology and an AHP-Weighted Information Content Method: A Case Study of Southern Anhui, China. ISPRS International Journal of Geo-Information. **8** (6), 266 (2019).

[24] Alzarooni E., Ali T., Atabay S., Yilmaz A. G., Mortula Md. M., Fattah K. P., Khan Z. GIS-Based Identification of Locations in Water Distribution Networks Vulnerable to Leakage. Applied Sciences. **13** (8), 4692 (2023).

[25] Shu H., Guo Z., Qi S., Song D., Pourghasemi H., Ma J. Integrating Landslide Typology with Weighted Frequency Ratio Model for Landslide Susceptibility Mapping: A Case Study from Lanzhou City of Northwestern China. Remote Sensing. **13** (18), 3623 (2021).

[26] Mohammadi A., Kiani B., Mahmoudzadeh H., Bergquist R. Pedestrian Road Traffic Accidents in Metropolitan Areas: GIS-Based Prediction Modelling of Cases in Mashhad, Iran. Sustainability. **15** (13), 10576 (2023).

[27] Salomão Í. L., Pinheiro P. R. Exploring Analytical Hierarchy Process for Multicriteria Assessment of Reinforced Concrete Slabs. Applied Sciences. **13** (17), 9604 (2023).

[28] Faul F., Erdfelder E., Buchner A., Lang A.-G. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. Behavior Research Methods. **41** (4), 1149–1160 (2009).

[29] Guan Y., Fu G.-H. A Double-Penalized Estimator to Combat Separation and Multicollinearity in Logistic Regression. Mathematics. **10** (20), 3824 (2022).

[30] Stoltzfus J. C. Logistic Regression: A Brief Primer. Academic Emergency Medicine. **18** (10), 1099–1104 (2011).

[31] Sainani K. L. Logistic Regression. PM&R. **6** (12), 1157–1162 (2014).

[32] Sperandei S. Understanding logistic regression analysis. Biochemia Medica. **24** (1), 12–18 (2014).

[33] Abalo R., Vernetta M., Gutiérrez-Sánchez A. Prevention of injuries to lower limbs using logistic regression equations in aerobic gymnastics. Medicina Dello Sport. **66** (2), 265–276 (2013).

[34] Lapresa D., Arana J., Anguera M. T., Pérez-Castellanos J. I., Amatria M. Application of logistic regression models in observational methodology: Game formats in grassroots football in initiation into football. Anales de Psicologia. **32** (1), 288–294 (2016).

[35] Shipe M. E., Deppen S. A., Farjah F., Grogan E. L. Developing prediction models for clinical use using logistic regression: an overview. Journal of Thoracic Disease. **11** (Suppl 4), S574–S584 (2019).

[36] Junus A., Hsu Y.-C., Wong C., Yip P. S. F. Is internet gaming disorder associated with suicidal behaviors among the younger generation? Multiple logistic regressions on a large-scale purposive sampling survey. Journal of Psychiatric Research. **161**, 2–9 (2023).

[37] Lavanya K., Rambabu P., Suresh G. V., Bhandari R. Gene expression data classification with robust sparse logistic regression using fused regularisation. International Journal of Ad Hoc and Ubiquitous Computing. **42** (4), 281–291 (2023).

[38] Seitshiro M. B., Mashele H. P. Assessment of model risk due to the use of an inappropriate parameter estimator. Cogent Economics & Finance. **8** (1), 1710970 (2020).

[39] Lecun Y., Bottou L., Bengio Y., Haffner P. Gradient-based learning applied to document recognition. Proceedings of the IEEE. **86** (11), 2278–2324 (1998).

[40] Zou D., Cao Y., Zhou D., Gu Q. Gradient descent optimizes over-parameterized deep ReLU networks. Machine Learning. **109** (3), 467–492 (2020).

[41] Chapelle O., Vapnik V., Bousquet O., Mukherjee S. Choosing multiple parameters for support vector machines. Machine Learning. **46** (1/3), 131–159 (2002).

[42] Saaty T. L., Kearns K. P. The Analytic Hierarchy Process. Analytical Planning. 19–62 (1985).

[43] Costa W. S., Pinheiro P. R., dos Santos N. M., Cabral L. D. A. F. Aligning the Goals Hybrid Model for the Diagnosis of Mental Health Quality. Sustainability. **15** (7), 5938 (2023).

[44] Saha A., Mandal S., Saha S. Geo-spatial approach-based landslide susceptibility mapping using analytical hierarchical process, frequency ratio, logistic regression and their ensemble methods. SN Applied Sciences. **2** (10), 1647 (2020).

[45] Hu X., Si M., Luo H., Guo M., Wang J. The Method and Model of Ecological Technology Evaluation. Sustainability. **11** (3), 886 (2019).

[46] Chen H., Chen J., Ding J. Data Evaluation and Enhancement for Quality Improvement of Machine Learning. IEEE Transactions on Reliability. **70** (2), 831–847 (2021).

[47] Vela D., Sharp A., Zhang R., Nguyen T., Hoang A., Pianykh O. S. Temporal quality degradation in AI models. Scientific Reports. **12** (1), 11654 (2022).

[48] Tay J. K., Narasimhan B., Hastie T. Elastic Net Regularization Paths for All Generalized Linear Models. Journal of Statistical Software. **106** (1), (2023).

[49] Sghir N., Adadi A., Lahmer M. Recent advances in Predictive Learning Analytics: A decade systematic review (2012–2022). Education and Information Technologies. **28** (7), 8299–8333 (2023).

[50] Baker S. G., Schuit E., Steyerberg E. W., Pencina M. J., Vickers A., Moons K. G. M., Mol B. W. J., Lindeman K. S. How to interpret a small increase in AUC with an additional risk prediction marker: decision analysis comes through. Statistics in Medicine. **33** (22), 3946–3959 (2014).

[51] Bejani M. M., Ghatee M. A systematic review on overfitting control in shallow and deep neural networks. Artificial Intelligence Review. **54** (8), 6391–6438 (2021).

[52] Chan C. P., Yang J. H., Chang W.-H. Entropy-based Time-series Financial Distress Model Based on Attribute Selection and MetaCost Methods for Imbalance Class. Proceedings of the 2023 3rd International Conference on Artificial Intelligence, Automation and Algorithms. 140–151 (2023).

[53] Stanlly, Putra F. A., Qomariyah N. N. DOTA 2 Win Loss Prediction from Item and Hero Data with Machine Learning. 2022 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT). 204–209 (2022).

[54] Lyu S., Zhao N., Zhang Y., Chen W., Zhou H., Zhu T. Predicting Risk Propensity Through Player Behavior in DOTA 2: A Cross-Sectional Study. Frontiers in Psychology. **13**, 827008 (2022).

[55] Trivedi U. B., Bhatt M., Srivastava P. Prevent Overfitting Problem in Machine Learning: A Case Focus on Linear Regression and Logistics Regression. Innovations in Information and Communication Technologies (IICT-2020). Advances in Science, Technology & Innovation. 345–349 (2021).

Покращення моделі логістичної регресії за допомогою оптимізації ваги, ініціалізованої AHP, із застосуванням регулярізації та адаптації градієнтного спуску: порівняльне дослідження

Камал А. А. [1,2], Мансор М. А. [1], Касіхмуддін М. С. М.[3]

[1] *Школа дистанційної освіти, Університет Сайнс Малайзія, 11800 USM, Малайзія*
[2] *Центр доуніверситетських досліджень, Університет Малайзії Саравак, 94300 Кота Самарахан, Малайзія*
[3] *Школа математичних наук, Університет Сайнс Малайзія, 11800 USM, Малайзія*

У цій статті досліджується підхід до покращення продуктивності моделі логістичної регресії (LR), інтегрованої з аналітичним ієрархічним процесом (AHP) для моделі ініціалізації ваги з регуляризацією та адаптацією градієнтного спуску (GD). Традиційна модель LR покладається на випадкову ініціалізацію ваги, що призводить до неоптимальних характеристик. Використовуючи AHP, отримано гібридну модель, яка використовує пріоритетний вектор як початкові ваги, що відображає відносну важливість вхідних характеристик. Попередні роботи повідомляли про низькі характеристики гібридної моделі AHP-LR через відсутність оптимізації для ініціалізованих ваг. У цьому дослідженні пропонується оптимізувати вагові коефіцієнти за допомогою підходу регуляризації L1 та L2, штрафуючи відхилення від ініціалізованих вагових коефіцієнтів AHP за допомогою модифікованої функції логарифмічної правдоподібності з модифікованою оптимізацією GD. Цей порівняльний аналіз включає чотири моделі: LR з регуляризацією L2, ваги AHP як ваги LR та ваги AHP, оптимізовані з регуляризацією L1 та L2. Прогнозний експеримент проводиться з використанням синтетичного набору даних для оцінки продуктивності моделей щодо точності, повноти, прецизійності, оцінки F1 та ROC-AUC. Результати показують, що оптимізація вагових коефіцієнтів з регуляризацією L1 або L2 значно покращує продуктивність моделі порівняно з прямим застосуванням вагових коефіцієнтів AHP без оптимізації, що дає майже випадкові припущення. Крім того, включення справжніх експертних вагових коефіцієнтів, оцінка їхнього впливу на продуктивність моделі та експериментування з автентичним набором даних і різними методами визначення вагових коефіцієнтів дадуть цінну інформацію.

Ключові слова: *логістична регресія; аналітичний ієрархічний процес; гібридна модель; регулярізація; модифікований градієнтний спуск.*