Volume 2, Issue 1, 2025

Print ISSN: 3007-6897

Online ISSN: 3007-6900

Multidisciplinary Journal of Engineering and Technology

Copyright© Upubscience Publisher

Multidisciplinary Journal of Engineering and Technology

Volume 2, Issue 1, 2025

Published by Upubscience Publisher

Copyright© The Authors

Upubscience Publisher adheres to the principles of Creative Commons, meaning that we do not claim copyright of the work we publish. We only ask people using one of our publications to respect the integrity of the work and to refer to the original location, title and author(s).

Copyright on any article is retained by the author(s) under the Creative Commons
Attribution license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Authors grant us a license to publish the article and identify us as the original publisher.

Authors also grant any third party the right to use, distribute and reproduce the article in any medium, provided the original work is properly cited.

Multidisciplinary Journal of Engineering and Technology
Print ISSN: 3007-6897 Online ISSN: 3007-6900
Email: info@upubscience.com
Website: http://www.upubscience.com/

Table of Content

Multidisciplinary Journal of Engineering and Technology
ISSN: 3007-6900
DOI: https://doi.org/10.61784/mjet3019

EARLY BEARING FAULT DETECTION AND RECOGNITION METHOD BASED ON INSTANCE TRANSFER

Zhao Jiang, Yu Wei*
School of International Business, Zhejiang Yuexiu University, Shaoxing 312000, Zhejiang, China.
Corresponding Author:Yu Wei, Email: jzgood@163.com

Abstract: Under the condition of data category imbalance, this paper proposes a TrAdaBoost-Least Squares Support Vector Machine (TrAdaBoost-LSSVM) algorithm based on instance transfer to solve the problem of low diagnostic accuracy of traditional machine learning methods. Firstly, use the K-means algorithm to screen the source domain data in order to eliminate those data with low similarity to the target domain, and then increase the inter-domain similarity; then, optimize the evaluation index of the base classifier to improve the model generalization ability. The simulation test results show that the method proposed in this paper exhibits the advantage of high fault recognition accuracy compared with the traditional machine learning method.
Keywords: Data category imbalance; Instance transfer; Machine learning; Least squares support vector machine

1 INTRODUCTION

With the rapid development of sensing technology and machine learning technology, data-driven intelligent fault detection and identification methods have received widespread attention [1-2]. Data-driven intelligent diagnostic algorithms are based on the assumption that the amount of sample data in each category is basically balanced, and the classification of unbalanced data in practical applications is widely used in product and equipment fault diagnosis [3-4]. Since the normal operation time of mechanical equipment is much larger than the time of equipment fault state, the fault state data of mechanical equipment is difficult to obtain, and the number of fault samples is usually much smaller than the number of samples under normal conditions [5-7]. For the sample distribution situation of category imbalance, the performance of traditional intelligent diagnosis may be seriously reduced, and even lose the actual diagnostic significance [8-9]. Therefore, there is an urgent need to improve the generalization of rolling bearing fault identification based on the category imbalance fault state recognition technique.

At present, the research on classifying category-imbalanced data mainly focuses on the data level and the algorithm level. At the data level, over-sampling and under-sampling are mainly used [10-11], but the shortcomings are that the over-sampling method introduces noise and the under-sampling method loses important information. At the algorithmic level, integration learning methods are mainly used to train different base classifiers for the same training set, and then integrate the base classifiers to form a stronger final classifier. Among them, the TrAdaBoost algorithm proposed by Wang et al. and the Bagging algorithm proposed by Freund et al. are more classic [12-13], and the TrAdaBoost algorithm is a kind of iterative algorithm, and the modified weights of the classifiers can be used as the basis for the final classifier. TrAdaBoost algorithm is an iterative algorithm, in which a new dataset with modified weights is fed into the next base classifiers for training, and then the base classifiers are integrated as the final decision classifiers. TrAdaBoost algorithm is advantageous in that it is able to complement the strengths of a number of base classifiers, which can improve the accuracy of the classifiers significantly. TrAdaBoost also has a problem of mismatch of the weights. Xia et al. show that assigning more initial weights to the target samples can alleviate this problem [14]; Wilson et al. show that increasing the proportion of classifiers for the more difficult samples reduces the proportion of classifiers for the majority of the samples [15]; and Hao et al. propose to resample the data in each iteration [16].

In view of the shortcomings of data-driven intelligent diagnostic algorithms in the classification method when dealing with category-imbalanced data, this paper proposes an improved TrAdaBoost method based on instance transfer. The method first uses a clustering algorithm to filter the source domain data to improve the similarity of data distribution between domains. Then the evaluation index of the base classifier is optimized to improve the generalization ability of the model. Finally, the feasibility of the proposed method is verified by the experimental data of accelerated life of rolling bearings from Xi'an Jiaotong University.

2 DYNAMIC MODELING OF ROLLING BEARING HEALTH STATE

The harsh working conditions cause rolling bearing vibration signals to be often interfered by noise, which seriously affects the extraction of weak fault characteristics. The Local Mean Decomposition (LMD) method can separate the effective information from the noise for multi-scale analysis [17-19] . Meanwhile, the TrAdaBoost method based on instance transfer improves the important role of diagnostic accuracy under unbalanced dataset conditions.

2.1 Local Mean Decomposition

Noting that the vibration signal acquired by monitoring is, the $x(t), t = 1, 2, \cdots, L$, t is the sampling time, the L is the total

© By the Author(s) 2025, under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0).

sampling time. The LMD method is implemented as follows.
(1) Apply the sliding average method of smoothing to obtain the local mean function, the $m_{11}(t)$ and the envelope estimator $\chi_{11}(t)$ 。
(2) By placing the local mean function, the $m_{11}(t)$ Separate it out and get a pure FM signal $s_{11}(t)$ 。

$$h_{11}(t) = x(t) - m_{11}(t) \tag{1}$$

$$s_{11}(t) = h_{11}(t) / \chi_{11}(t) \tag{2}$$

Repeat until,is satisfied that $\lim_{n\to\infty} \chi_{1n}(t) = 1$ conditions.
(3) Multiply the envelope function to obtain the envelope signal, the

$$\chi_1(t) = \chi_{11}(t) \cdot \chi_{12}(t) \cdots \chi_{1n}(t) \tag{3}$$

(4) FM signals $s_{1n}(t)$ with the envelope signal $\chi_1(t)$ The Product Functions (PF) component is obtained by multiplying the Product Functions (PF).

$$PF_1(t) = s_{1n}(t) \cdot \chi_1(t) \tag{4}$$

(5) Will $PF_1(t)$ from $x(t)$ The new signal is obtained by separating it from the $\mu_1(t)$ Repeat q Times, until, uh $\mu_q(t)$ stops iterating when it is a monotonic function.

$$\begin{cases} \mu_1(t) = x(t) - PF_1(t) \\ \mu_2(t) = \mu_1(t) - PF_2(t) \\ \quad \vdots \\ \mu_q(t) = \mu_{q-1}(t) - PF_q(t) \end{cases} \tag{5}$$

(6) The vibration signal is finally decomposed into a series of PF components and a residual component, the $\mu_q(t)$ 。

$$x(t) = \sum_{i=1}^{q} PF_i(t) + \mu_q(t) \tag{6}$$

2.2 Extraction of Fault Features

In this paper, the LMD decomposition obtained by q The PF components were extracted in the following steps.
(1) Constructing a signaling matrix that $M(t) = \left[f_1(t), f_2(t), \cdots, f_q(t)\right]$;
(2) Calculate the covariance matrix, the H :

$$H = \frac{MM^T}{trace\left(MM^T\right)} \tag{7}$$

Where. $trace(\cdot)$ for the trace operation on the matrix inside the parentheses.
(3) For the matrix H Perform eigenvalue decomposition: the

$$H = U \Sigma U \tag{8}$$

Where. U for H matrix composed of eigenvectors; Σ is the diagonal matrix composed of the corresponding eigenvalues.
(4) Use it to whiten the covariance matrix to obtain the matrix, the S :

$$\begin{cases} p = \sqrt{\Sigma^{-1}} U^T \\ S = pHp^T \end{cases} \tag{9}$$

(5) Perform the eigenvalue decomposition and construct the space that $\tilde{X}$

$$\begin{cases} S = U_1 \Sigma_1 U_1^T \\ \tilde{X} = U_1^T p \end{cases} \tag{10}$$

Where. U_1 is the matrix consisting of S eigenvectors; the Σ_1 is the diagonal matrix formed by the corresponding eigenvalues.
(6) Perform the projection to obtain the principal element space and the residual space: the

$$\bar{X} = \tilde{X} + e \tag{11}$$

T^2 Control charts are a quality control tool based on the chi-square distribution for statistical process control. This control chart is mainly used to monitor the variability of categorical data, and to determine whether anomalies occur in the collected sample data by comparing the statistics with the control limits. The basis for anomaly detection is mainly the comparison of T^2 and Squared Prediction Error (SPE) statistics and control limits.
T^2 The statistic is used to measure the distance of the sample from the origin of the principal element space, the statistic expression is.

$$T^2 = U^T p \Lambda^{-1} p^T U \tag{12}$$

T^2 The expression for the control limit of

$$UCL = \frac{a\left(n^2-1\right)}{n\left(n-a\right)} F_{a(a,n-a)} \tag{13}$$

The SPE statistic is used to characterize the variance of the data and is obtained from the residual space with the expression.

$$SPE = \left\| \left(I - pp^T \right) U \right\|^2 \tag{14}$$

The control limit UCL of the SPE can be obtained from the probability density function, expressed as follows.

$$UCI = \theta_1 [\frac{h_0 C_a \sqrt{2\theta_2}}{\theta_1} + \frac{\theta_2 h_0 \left(h_0 - 1\right)}{\theta_1} + 1]^{\frac{1}{h_0}} \tag{15}$$

$$\theta_j = \sum_{i=t+1}^{T} \lambda_i^j , j = 1,2,3 \tag{16}$$

$$h_0 = 1 - \frac{2\theta_1 \theta_3}{3\theta_2} \tag{17}$$

In the formula, the C_a is normally distributed α Loci.

3 IMPROVED TRADABOOST ALGORITHM BASED ON INSTANCE TRANSFER

Sample data in different domains are similar, so the useful information in the source domain can be utilized to build classifiers in the target domain.TrAdaBoost has an advantage over Bagging algorithm in dealing with the problem of imbalance in the classification of datasets in that it is able to improve the accuracy of the classifier by adjusting the weights in real time to make a number of weak classifiers into a strong classifier [20-22] .

3.1 Screening for Redundant Data

The process of filtering redundant data is as follows.
Input: source domain dataset is $T_a = \{(x_a^i, y_a^i)\}_{i=1}^m$, target domain dataset $T_b = \{(x_b^j, y_b^j)\}_{j=1}^n$, the number of data blocks is n 。

Output: Filtered source domain dataset $T_a^{'}$ and target domain datasets $T_b^{'}$ 。

(1) Construct the training dataset that $T = T_a \cup T_b$ 。
(2) The dataset T is partitioned into n block of data.
(3) The K-means algorithm is used on the data block to filter out the data with little similarity to the target domain and remove them. The
(4) According to the source domain feature distribution after screening, select the features that are closest to the target domain feature distribution.

3.2 Steps to Improve the Implementation of the TrAdaBoost Algorithm Based on Instance Transfer

TrAdaBoost algorithm and Bagging have obvious shortcomings: in the iterative process of the error samples are given a larger weight, and the correct samples are given a smaller weight, which can easily lead to an infinite increase in the weight of the error samples, and the correct samples are neglected, which affects the training effect of the samples [23-25]. Based on this deficiency, the TrAdaBoost algorithm is improved, and the improved algorithm process is as follows.
Input: source domain training set $D_{a_1}, D_{a_2}, ..., D_{a_N} = \{(x_{a_1}, y_{a_1}), (x_{a_2}, y_{a_2}), ..., (x_{a_N}, y_{a_N})\}$, the target domain training set $D_b = \{(x_b, y_b)\}$, the maximum number of transfers is N 。
Output: target classifier function that

$$F(x_i) = sign\{\prod_{h=1}^{\frac{N}{2}} [\beta_h^{-f_h(x_i)}] - \prod_{h=1}^{\frac{N}{2}} [(\beta_h)^{-1/2}] \tag{18}$$

(1) Preprocess the training dataset, call the K-means algorithm to output the filtered source domain dataset $T_a^{'}$ and target domain datasets $T_b^{'}$, the number of datasets are, respectively r and q 。
(2) The weight vector is initialized as $W^1 = (\omega_1^1, ..., \omega_{r+q}^1)$, where

$$\omega_i^1 = \begin{cases} \frac{1}{r} & i = 1,2,...,r \\ \frac{1}{q} & i = r+1,...,r+q \end{cases} \tag{19}$$

(3) Setting the source domain weight update factor, the

$$\beta = \frac{1}{1 + \sqrt{\frac{2\ln r}{N}}} \tag{20}$$

(4) Normalize the weights by

$$P^h = \frac{W^h}{\sum_{i=1}^{r+q} \omega_i^h} \tag{21}$$

Among them. $h = 1,...,N$

(5) For a weight distribution of P^h of the dataset $T^{'}$ training, input to LSSVM weak classifier, to get strong classifier f_b^h 。

(6) Calculation f_b^h exist $T_b^{'}$ The error rate on the

$$\pi_i^h = \frac{\sum_{i=r+1}^{r+q} \omega_i^h \left| y_b^j - f_b^h x_b^j \right|}{\sum_{i=r+1}^{r+q} \omega_i^h} \tag{22}$$

Among them. x_i for data set no i Data.

(7) Setting the target domain weight update factor, the

$$\beta_i^h = \pi_i^h / (1 - \pi_i^h) \tag{23}$$

(8) Updating the weight vector, the

$$\omega_i^{h+1} = \begin{cases} \omega_i^h \beta^{\left| y_b^j - f_b^h x_b^j \right|}, & 当 i = 1,2,...,r \\ \omega_i^h \beta_i^{-\left| y_b^j - f_b^h x_b^j \right|}, & 当 i = r+1,...r+q \end{cases} \tag{24}$$

if $\pi_i^h = 0$, then $\left| y_b^j - f_b^h x_b^j \right|$ is denoted as a sample x_b^j . The probability of belonging to category 0 but being misclassified as category 1. Similarly, if the $\pi_i^h = 1$, then $\left| y_b^j - f_b^h x_b^j \right|$ is denoted as a sample x_b^j. Probability of belonging to category 1 but being misclassified as category 0.

4 ROLLING BEARING FAULT DETECTION AND ABNORMAL OPERATION STATE IDENTIFICATION METHOD FLOW

4.1 Based on LMD-PCA Rolling Bearing Fault Detection

Rolling bearings are often interfered by noise due to complex and severe working conditions, which is not favorable to the extraction of weak fault characteristics [26-29]. In this paper, the LMD method is used to separate the effective information from the noise, and on the basis of which the PCA method is applied to extract the fault features. Based on the LMD-PCA rolling bearing fault detection framework, which contains the following three main steps.

(1) Dynamic modeling of vibration signals. The vibration signal is decomposed using the LMD method, and is obtained by n A matrix composed of the IMF components of the X ;

(2) For the matrix, the X Standardization;

(3) The PCA method is used to decompose the eigenvalues of the matrix to obtain the principal element space and the residual space. Then the main element space and residual space are decomposed several times to obtain the subspace, and the statistics and control limits of each subspace are calculated and compared to determine whether there is a fault.

(4) Abnormal operation state identification.

4.2 Improved TrAdaBoost Anomalous Run State Identification Method Based on Instance Transfer

The schematic diagram of the improved TrAdaBoost multi-classification algorithm.The steps are as follows.

(1) Segmentation of source and target domain samples; and

(2) Setting up the training and test sets;

(3) initialize the weights $W_{n+m} = (1/n,...,1/n,1/m,...,1/m)$, samples are drawn from the joint training set to obtain the subsample set;

(4) Update the weights using the error on the multiple classifiers to obtain a new set of subsamples;

(5) Repeat (3) to (4) N times, to obtain the abnormal operation state recognition model $F_N(x_i)$;

(6) Repeat (5) Z times, to obtain Z abnormal operation state identification models $F_{Nj}(x_i)$, $j = 1,2,...,Z$;

(7) Category-consistent voting on the Z results.

5 EXAMPLE ANALYSIS

The fault detection model should have good accuracy and real-time performance. Firstly, the LMD-PCA method is used to extract the features, and the statistics are calculated sequence by sequence according to the sliding window, and compared with the control limit to determine whether the bearings are abnormal or not. Secondly, when the abnormal state of the bearing is detected, the TrAdaBoost algorithm is used to recognize the abnormal operation state based on the instance transfer improvement.

5.1 Experimental Data Collection and Parameterization

The data set selected for this experiment is the XJTU-S data set [30] . The experimental platform for collecting this dataset is shown in Fig. 3, which is capable of conducting accelerated degradation experiments on bearings, providing the actual degradation data of five rolling bearings modeled as LDKUER204 throughout their service life, as shown in Table 1 Two PCB352C33 accelerometers were placed on the vertical and horizontal axes to test the vibration signals. The data types and labels of XJTU-S dataset are shown in Table 1, the sampling frequency is set to 25600Hz, the loads of two working conditions are 11kN and 12kN, and the driving speeds of the control motor are 2100r/min and 2400r/min, respectively.The types of faults include rolling body faults, inner ring faults, outer ring faults and cage faults, and there are four types of samples.In order to validate the effectiveness of the proposed In order to verify the validity of the proposed method to recognize the fault state of rolling bearings, 1000 sampling points are taken as a sample.

Table 1 Data Types and Labels of the XJTU-SY Dataset

data sets	Sample length	the type of fault	data sets	Sample length	the type of fault
Bearing 1_1	1000	inner ring failure	Bearing 2_1	1000	Rolling body failure
Bearing 1_2	1000	outer ring failure	Bearing 2_2	1000	Cage failure
Bearing 1_3	1000	Cage failure	Bearing 2_3	1000	inner ring failure
Bearing 1_4	1000	outer ring failure	Bearing 2_4	1000	inner ring failure
Bearing 1_5	1000	Rolling body failure	Bearing 2_5	1000	outer ring failure

5.2 Rolling Bearing Fault Detection

In order to verify the effectiveness of the LMD-PCA method proposed in this paper for the fault detection of the outer ring of rolling bearings, the vibration signals of the bearing 2_4 inner ring fault are taken as an example, firstly, the LMD decomposition is carried out and then reconstructed, and then the features of the bearing are extracted using the principal component analysis (PCA) method, and by comparing it with the traditionalempirical modal decompositionThe Empirical Mode Decomposition (EMD)-PCA method is compared to verify the detection capability of the method for the early faults of rolling bearings. Figure 1 depicts the fault detection results of the inner ring of rolling bearing based on the traditional EMD-PCA method, where the red dotted lines indicate the control limits and the black solid lines indicate the statistics, and Figure 2 depicts the fault detection results of the outer ring of rolling bearing based on the LMD-PCA method. The red dotted line indicates the control limit, and the black solid line indicates the statistics. From the Figure 2, it can be concluded that the LMD-PCA method can detect the abnormal state earlier than the traditional EMD-PCA method.

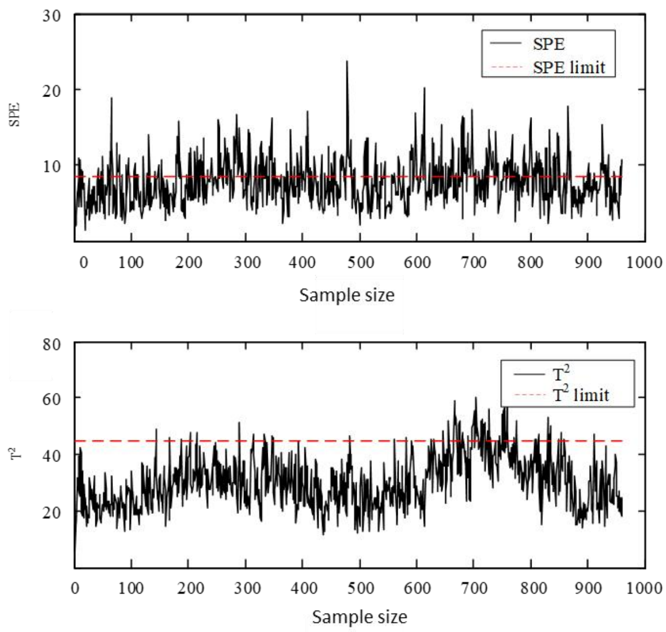


Figure 1 Detection Results of the EMD-PCA Method

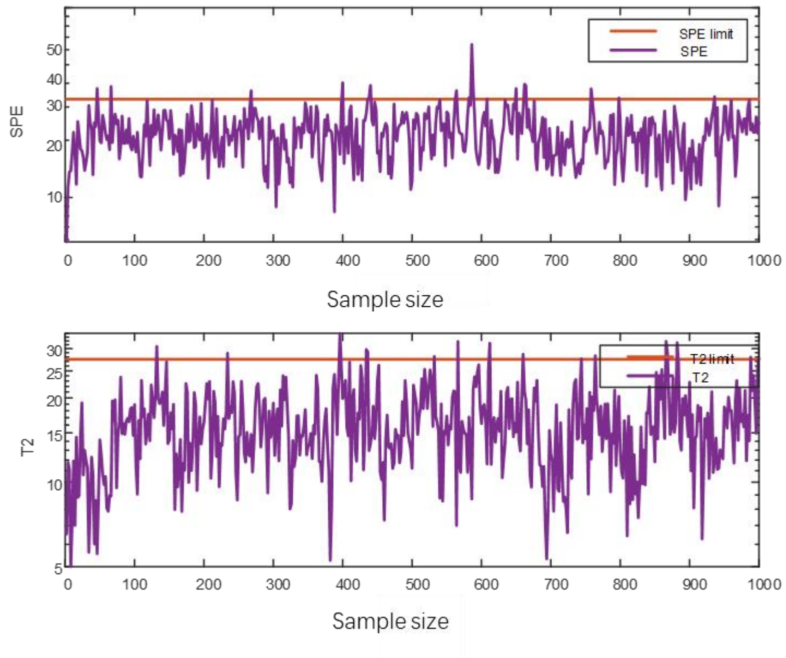


Figure 2 Detection Results of the LMD-PCA Method

5.3 Rolling Bearing Abnormal State Identification

In order to verify the validity of the method in this paper, the following experiments are designed with 11kN load and 2100r/minr motor drive speed as the auxiliary samples in the source domain, and 12kN load and 2400r/min motor drive speed as the samples in the target domain: 1. Training set: 400 normal state training samples in the source domain, and 40, 80, 120, 160 (10, 20, 30, 40 for each of the four fault states) training samples in the target domain; 2. Test set: 240 unlabeled samples (10, 20, 30, 40 for each of the four fault states) in the target domain; and 3. 160 samples (10,20,30,40 samples for each of the 4 fault states); 2. Test set: 240 unlabeled samples in the target domain (60 samples for each of the 4 fault states), and 100 normal state samples in the source domain to form a class-imbalanced distribution.

LSSVM was used as the base classifier of the algorithm and optimization was carried out using the Genetic Algorithm Toolbox of the University of Sheffield, UK.The average recognition rate of faulty operating states was used as the fitness function to obtain the parameter combinations of the LSSVM as. $\iota = 0.71, \gamma = 5.56, g = 7.75$. Finally, the recognition rate of faulty operation state is obtained by calculating and comparing the performance of two algorithms: (1) SVM method, and (2) LSSVM classifier with optimized parameters is used as a weak classifier of AdaBoost. The results of the experiment are shown in Table 2.

Table 2 Faulty Operating State Recognition Rate (%)

Source domain samples	Target domain training samples	SVM	AdaBoost-LSSVM	TrAdaBoost-LSSVM
	40	65.6	65.3	69.5
400	80	73.4	77.6	79.8
	120	81.2	89.6	92.5
	160	89.6	94.8	95.7

5.4 Discussion of N-values vs. Z-values

Consider whether different ratios of the number of data samples in the target domain of the test set to the number of data samples in the source domain would have an impact on the accuracy of the algorithm,setting $N = 10, Z = 1000$,The percentage of randomly selected target domain data was set to 5%, 10%, 15%, 20%, 25%, and 30% in turn. The experimental results are shown in Table 3.

Table 3 Accuracy of the Algorithm for Different Proportions of Target Domain Data

Methods	5%	10%	15%	20%	25%	30%
TrAdaBoost-LS-SVM	0.66	0.69	0.72	0.79	0.87	0.92
SVM	0.53	0.65	0.67	0.73	0.75	0.81
AdaBoost-LSSVM	0.56	0.66	0.69	0.76	0.78	0.89

As can be seen from Table 3, the classification accuracy of the TrAdaBoost-LSSVM algorithm based on instance transfer is significantly higher than that of the other algorithms, indicating that after the increase of training data in the target domain, the inter-domain similarity is getting higher and higher, and the classification performance of the classifier is better.

For the values of N and Z, first set Z to a larger value ($Z = 1000$) Mr. José Antonio González, Minister Counsellor, Permanent Mission $N \in [1, 20]$ corresponding to the recognition rate at the time.

As shown in Fig. 6,when, the $N \in [1,12]$ .When ,the recognition rate of each group of samples varies with N value increases; when, the N When it exceeds 15, the recognition rate increases N The increasing of the decreases instead. Take the $N \in [12,14]$ is reasonable.

Although $Z = 1000$ Ensure the stability of diagnostic results, but the algorithm is more time-consuming. when the Z value is 500 and above, the variance of the recognition rate is already smaller and begins to stabilize. Therefore, the $Z = 500$ It can already meet the requirements for use.

6 CONCLUSION

In the case of complex working conditions and lack of training sample data, the stability of online detection is very important, and the reduction of false alarm rate is an important goal to improve the effectiveness of online detection. Therefore, in this paper, we extract the fault characteristics by LMD method, calculate the statistics of bearing data in different working conditions and determine the control limit as the threshold standard to detect whether the abnormal state occurs or not. Secondly, the TrAdaBoost-LSSVM algorithm based on instance transfer is used to recognize the abnormal state of the bearing. The method firstly uses the K-means algorithm to filter the source domain data, eliminate those data with low similarity to the target domain, and then increase the inter-domain similarity. Then the evaluation index of the base classifier is optimized to improve the generalization ability of the model. The experimental results prove the feasibility and effectiveness of the proposed method, and enhance the ability of recognizing the fault features of weak samples.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

FUNGING

The research achievements of the scientific research project "Research on the Mechanism of Intelligent Assessment and Remaining Useful Life Prediction of Lithium-Ion Battery State Based on Transfer Learning" at Zhejiang Yuexiu University of Foreign Languages in 2024 (Project Number: N2024011).

REFERENCES

[1] WANG Taiyong, GONG Liming, WANG Peng, et al. Fault diagnosis model of rotating machinery based on KD -DenseNet. Vibration and Shock, 2020, 39(16): 39-45.

[2] Xingxing Zhang, Shaobo Li, Longxuan Tsu, et al. Research on rolling bearing fault diagnosis based on machine learning algorithm. Combined machine tools and automated machining technology, 2020, 10(7): 36 -39.

[3] CHEN Renxiang, HUANG Xin, YANG Lixia, et al. Rolling bearing fault diagnosis based on convolutional neural network and discrete wavelet transform. Journal of Vibration Engineering, 2018, 31(5): 883-891.

[4] Mao WT, Tian SY, Dou Z, et al. A deep transfer learning-based online detection method for early faults in rolling bearings. Journal of Automation, 2022, 48(1): 302-314.

[5] Liu JC, Quan H, Yu X, et al. Rolling bearing fault diagnosis based on parameter optimization VMD and sample entropy. Journal of Automation, 2022, 48(3): 808-819.

[6] Shakya P, Kulkarni M S, Darpe A K. A novel methodology for online detection of bearing health status for naturally progressing defect. Journal of Sound & Vibration, 2014, 333(21): 5614-5629.

[7] Taylor M E, Stone P H. Transfer learning for reinforcement learning domains: a survey. The Journal of Machine Learning Research, 2009, 10(10):1633-1685.

[8] Yang S, Liu Z, Lu G. Early change detection in dynamical bearing degradation process based on hierarchical graph model and adaptive inputs weighting fusion. IEEE Transactions on Industrial Informatics, 2021, 17(5): 3186-3196.
[9] XU Le, YU Ruxin, XING Bangsheng, et al. Rolling bearing fault feature extraction based on LMD arrangement entropy. Mechanical Transmission, 2019, 43(1): 136-139.
[10] Chawla N V, Bowyer K W, Hall L O, et al. SMOTE: Synthetic minority over-sampling technique . Journal of Artificial Intelligence Research, 2011, 16(1): 321-357.
[11] Yen S J, Lee Y S. Cluster-based Under-sampling Approaches for Imbalanced Data Distributions. Expert Systems with Applications, 2006, 36(3): 5718-5727.
[12] Wang W, Wang C, Wang Z, et al. Abnormal detection technology of industrial control system based on transfer learning. Applied Mathematics and Computation, 2022, 412(1): 126-139.
[13] Freund Y. A decision-theoretic generalization of on-line learning and an application to boosting // Proceedings of the Second European Conference on Computational Learning Theory. Springer-Verlag, 1995.
[14] Qian Z, Hai G L, Yong Z, et al. Instance transfer learning with multisource dynamic trAdaBoost. The Scientific World Journal, 2014, 20(14): 1-8.
[15] Wang C, Yang A, Yuan A, et al. Diagnosis of wind turbine faults with transfer learning algorithms. Renewable Energy, 2021, 16(3): 2053-2067.
[16] Xia R, Zong C, Hu X, et al. Feature ensemble plus sample selection: domain adaptation for Sentiment Classification . IEEE Intelligent Systems, 2013, 28(3): 10-18.
[17] Wilson C, Check J H, Amui J, et al. Effect of multiple sources vs. single source of donor embryos on pregnancy and Implantation Rates per Transfer . Clinical and experimental obstetrics & gynecology, 2011, 38(4):324-325.
[18] Hao S. Kernel methods for transfer learning to avoid negative transfer . International Journal of Computing Science and Mathematics, 2016, 7(2):175-190.
[19] Wang T, Liu Z, Lu G, et al. Temporal-spatio graph based spectrum analysis for bearing fault detection and diagnosis. IEEE Transactions on Industrial Electronics, 2021, 68(3): 2598 -2607.
[20] Li Y, Xu M, Wang R, et al. A fault diagnosis scheme for rolling bearing based on local mean decomposition and improved multiscale fuzzy entropy.Journal of Sound and Vibration, 2016, 3(6): 277-299.
[21] Lei Y, He Z, Zi Y. EEMD method and WNN for fault diagnosis of locomotive roller bearings. Expert Systems with Applications, 2011, 38(6): 7334-7341.
[22] Fu C, Wu Z, Xue M, et al. Cross-domain decision making based on TrAdaBoost for diagnosis of breast lesions. Artificial Intelligence Review, 2023, 56(5): 3987-4017.
[23] Zheng L, Liu G, Yan C, et al. Improved TrAdaBoost and its application to transaction fraud detection. IEEE Transactions on Computational Social Systems, 2020, 7(5): 1304-1316.
[24] He H, Khoshelham K, Fraser C. A multiclass TrAdaBoost transfer learning algorithm for the classification of mobile lidar data. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 16(6): 118-127.
[25] Yong Z, Chengbin W. An indoor positioning system using Channel State Information based on TrAdaBoost Tranfer Learning //2021 4th International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE). IEEE, 2021: 1286-1293.
[26] Wang W, Wang C, Wang Z, et al. Abnormal detection technology of industrial control system based on transfer learning. Applied Mathematics and Computation, 2022, 4(12): 126-139.
[27] GUO Xiaoping, LIU Shiyang, LI Yuan. Research on multi-operating condition process fault detection method based on sparse residual distance. Journal of Automation, 2019, 45(3): 618-625.
[28] ZHAO Zhihong, YANG Shaopu. A sample entropy-based bearing fault diagnosis method. Vibration and Impact, 2012, 31(6): 136-140.
[29] Su W-S, Wang F-T, Zhang Z-X, et al. Application of EM noise reduction and spectral cliff method in early fault diagnosis of rolling bearings. Vibration and Shock, 2010, 29(3): 18-21.
[30] Lei Yaguo, Han Tianyu, Wang Biao, et al. Interpretation of XJTU-SY rolling bearing accelerated life test dataset. Journal of Mechanical Engineering, 2019, 55(16): 1-6.

Multidisciplinary Journal of Engineering and Technology
ISSN: 3007-6900
DOI: https://doi.org/10.61784/mjet3018

DIGITAL FORENSIC INVESTIGATION IN CYBERCRIME CASES: CASE STUDIES AND RECOMMENDATIONS

Cut Rafifah Syaakirah, Luthfiyah Syifa, Iskandar Muda*
Department Accounting, Universitas Sumatera Utara, Medan, Indonesia.
Corresponding author: Iskandar Muda, Email: iskandar1@Student.usu.ac.id

Abstract: The exponential advancement of technology and widespread internet adoption have introduced numerous advantages while simultaneously giving rise to an alarming surge in cybercrime activities such as hacking, financial fraud, identity theft, and data breaches. Digital forensic investigation has emerged as a pivotal mechanism in addressing these crimes by identifying offenders, gathering actionable evidence, and aiding judicial processes. This paper delves into the utilization of digital forensic methods in resolving cybercrime incidents, drawing on a selection of illustrative case studies to showcase the field's strengths and limitations. The research begins with an overview of the foundational concepts and techniques in digital forensics, emphasizing its interdisciplinary approach that integrates computing, legal frameworks, and investigative practices. Detailed case studies illustrate practical applications, including tracing ransomware campaigns, uncovering insider breaches within corporate environments, and analyzing fraudulent cryptocurrency activities.
Keywords: Integrating insights; Malware deconstruction; Financial fraud

1 INTRODUCTION

The essential role of forensic tools, such as log analysis, network traffic examination, malware deconstruction, and data recovery techniques, in building robust investigative outcomes. Despite its potential, digital forensic investigations face significant obstacles, including the proliferation of encryption technologies, the diversity of digital devices and platforms, and the jurisdictional hurdles posed by transnational cybercrime [1]. To address these issues, this paper proposes actionable strategies, such as adopting uniform investigation protocols, enhancing forensic tools and training, fostering international law enforcement cooperation, and refining legal standards to ensure the reliability and admissibility of digital evidence. By integrating insights from case studies, scholarly literature, and industry practices, this paper underscores the evolving role of digital forensics as an indispensable tool in combating cybercrime. It advocates for continued innovation and cross-disciplinary collaboration to stay ahead of emerging threats and maintain the integrity of justice in an increasingly digital landscape.

2 LITERATURE REVIEW

Digital forensic investigation has emerged as a critical discipline in addressing the complex challenges posed by cybercrime. This field, which integrates principles from computer science, criminology, and law, focuses on the identification, preservation, analysis, and presentation of digital evidence [2]. The literature surrounding digital forensic investigation has grown substantially, reflecting the field's rapid evolution and increasing relevance in the face of escalating cyber threats. This review synthesizes key scholarly works, technical advancements, and practical applications relevant to the investigation of cybercrime.

2.1 The Foundations of Digital Forensics

Digital forensics has its roots in traditional forensic science but has evolved to address the unique challenges of digital evidence. Early studies emphasized the need for standardized processes, such as those outlined by [3] who introduced the concept of a "four-stage process" involving acquisition, examination, analysis, and reporting. This foundational framework continues to guide modern practices. Complementary research has underscored the importance of adhering to legal and ethical standards to ensure that digital evidence remains admissible in court.

2.2 Cybercrime and Its Impact

The growth of cybercrime, ranging from ransomware attacks and phishing scams to advanced persistent threats (APTs), has been well-documented in the literature. Symantec's annual reports and other industry analyses have shown a year-on-year increase in the sophistication and scale of cyber-attacks. Scholars like [4] have explored the economic impact of cybercrime, noting its significant cost to organizations and governments. These studies emphasize the urgency of developing robust digital forensic capabilities to counter these threats[5].

© By the Author(s) 2025, under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0).

2.3 Forensic Techniques and Tools

Numerous studies have explored the tools and techniques utilized in digital forensic investigations. Traditional methods such as file system analysis, memory imaging, and keyword searches remain widely used. However, advancements in technology have led to the development of specialized tools such as EnCase, FTK (Forensic Toolkit), and Volatility, which enable detailed analyses of complex systems. Research by [6]introduced "scalable forensics," which focuses on processing large datasets efficiently, an increasingly important capability given the growing volume of digital evidence.

Emerging technologies, such as artificial intelligence and machine learning, are also gaining traction in digital forensics. These technologies are being applied to automate repetitive tasks, identify patterns in large datasets, and enhance predictive capabilities. For example, [7] demonstrated how machine learning could be used to analyze network traffic and detect anomalies indicative of cybercrime activities.

2.4 Challenges in Digital Forensic Investigations

The literature identifies several challenges that complicate digital forensic investigations. One of the most significant is the increasing use of encryption, which can render digital evidence inaccessible. A study by [8] highlighted the prevalence of encryption in ransomware attacks, complicating efforts to recover victim data. Another challenge is the diversity of digital devices and platforms, ranging from mobile devices and IoT gadgets to cloud-based services. Research by Quick and Choo explored the complexities of cloud forensics, emphasizing the need for tools and techniques tailored to distributed environments.

Jurisdictional challenges also feature prominently in the literature, as cybercrime often transcends national boundaries. Authors such as [9] have explored the legal and procedural hurdles that arise when evidence and perpetrators are located in different countries. These challenges underscore the importance of international cooperation and standardized protocols.

2.5 Case Studies in Digital Forensics

Case studies play a vital role in understanding the practical application of digital forensic techniques. Notable examples include investigations into high-profile ransomware campaigns such as WannaCry and NotPetya, which have been extensively analyzed in both academic and industry publications. These cases highlight the effectiveness of digital forensics in tracing malware origins and identifying the perpetrators.

Another area of focus is insider threats, where employees misuse their access to organizational resources for malicious purposes. Studies by Hu et al. have demonstrated how forensic tools can be used to uncover suspicious activity, such as unauthorized data transfers or tampering with critical systems.

Cybercrime involving cryptocurrencies has also gained attention, with researchers like [10] examining forensic techniques for tracing Bitcoin transactions. These studies reveal the dual challenge of navigating pseudonymity and the decentralized nature of blockchain networks.

2.6 Recommendations from the Literature

The literature consistently emphasizes the need for a proactive and adaptive approach to digital forensic investigations. Key recommendations include the development of advanced tools capable of handling encrypted and cloud-based data, increased investment in practitioner training, and fostering stronger collaboration between academia, industry, and law enforcement. Additionally, studies advocate for the establishment of international agreements to facilitate cross-border investigations and streamline the sharing of digital evidence.

3 METHODOLOGY

This study employs a qualitative research methodology to explore the practices, challenges, and recommendations associated with digital forensic investigations in cybercrime cases. The qualitative approach is chosen due to its capacity to provide an in-depth understanding of complex phenomena, particularly in areas where human expertise, contextual interpretation, and subjective experiences are critical. By analyzing real-world cases, expert insights, and existing literature, this methodology aims to uncover patterns, generate meaningful interpretations, and propose actionable recommendations.

3.1 Research Design

The study adopts an exploratory research design to investigate how digital forensic techniques are applied in cybercrime investigations. The design emphasizes a case study approach, complemented by thematic analysis, to examine real-world instances where digital forensic methodologies were utilized. This combination allows for a detailed examination of specific scenarios while identifying recurring themes and challenges that transcend individual cases.

3.2 Data Collection Methods

3.2.1 Case study analysis
The study relies on documented case studies of cybercrime investigations, sourced from scholarly articles, industry reports, and forensic analysis publications. Selected cases include high-profile incidents such as ransomware attacks, insider data breaches, and cryptocurrency fraud. Each case study provides rich insights into the forensic methods applied, the challenges encountered, and the outcomes achieved.

3.2.2 Expert interviews
Semi-structured interviews are conducted with professionals in the field of digital forensics, including forensic analysts, cybersecurity experts, and law enforcement officials. The interviews aim to capture their firsthand experiences, perspectives on emerging challenges, and recommendations for improving forensic practices. Open-ended questions encourage detailed responses, allowing the researcher to probe further into specific areas of interest.

3.2.3 Document analysis
Additional data is gathered from secondary sources, including government reports, legal documents, and policy frameworks related to digital forensic practices. This data provides a broader understanding of the legal and procedural contexts within which digital forensics operates.

3.3 Sampling Strategy

3.3.1 Case selection
Cases are purposively selected based on their relevance to the study's objectives. The selection criteria include the type of cybercrime, the complexity of the forensic investigation, and the availability of detailed documentation. Priority is given to cases that illustrate diverse challenges and solutions in digital forensics.

3.3.2 Participant selection
Expert participants are identified through professional networks and industry affiliations. Purposive sampling is used to ensure the inclusion of individuals with significant experience and expertise in digital forensic investigations. Efforts are made to achieve diversity in terms of professional roles, sectors, and geographical regions to capture a wide range of perspectives.

3.4 Data Analysis

3.4.1 Thematic analysis
A thematic analysis is conducted to identify patterns and themes across the data collected from case studies, interviews, and documents. The process involves coding the data, categorizing codes into themes, and interpreting the relationships between themes. This approach allows the study to uncover common challenges, innovative practices, and areas for improvement in digital forensic investigations.

3.4.2 Cross-Case comparison
The case studies are compared to identify similarities and differences in the application of forensic techniques and the challenges encountered. This comparison helps to highlight best practices and contextual factors that influence the success of forensic investigations.

3.4.3 Triangulation
To enhance the credibility of the findings, data from different sources—case studies, interviews, and document analysis—are cross-referenced. Triangulation ensures that the conclusions drawn are robust and well-supported by evidence.

3.5 Ethical Considerations

The study adheres to ethical guidelines to ensure the integrity of the research process. Informed consent is obtained from all interview participants, and their anonymity is protected to ensure confidentiality. Data from publicly available case studies and documents are used responsibly, with proper attribution to original sources.

3.6 Limitations

While the qualitative methodology provides deep insights, it is inherently limited by its reliance on subjective interpretation and non-generalizable findings. The study’s focus on purposively selected cases and participants may also introduce a degree of selection bias. To address these limitations, the findings are framed within the specific context of the research and supplemented with broader literature to ensure relevance and applicability.

4 RESULTS

This section presents the findings of the study on digital forensic investigations in cybercrime cases. The results are derived from a detailed analysis of selected case studies, expert interviews, and relevant documents, highlighting the practices, challenges, and recommendations for improving digital forensic processes. These findings are organized into key themes, including the effectiveness of forensic techniques, the challenges encountered in investigations, and the emerging trends in the field.

4.1 Effectiveness of Digital Forensic Techniques

Digital forensic tools and methodologies were found to play a pivotal role in uncovering critical evidence, enabling investigators to reconstruct events, identify perpetrators, and support legal proceedings. Case studies demonstrated that:

4.1.1 File system analysis and data recovery

File system analysis remains a cornerstone of digital forensic investigations. In several cases, investigators successfully recovered deleted files and hidden data, providing essential evidence for cybercrime prosecutions. For instance, in a ransomware case, forensic experts used advanced recovery tools to retrieve encryption keys stored in system memory, allowing victims to regain access to their data.

4.1.2 Network traffic analysis

Analyzing network traffic proved invaluable in identifying the source of attacks and understanding the methods used by perpetrators. In a distributed denial-of-service (DDoS) attack case, forensic analysts used packet capture tools to trace malicious traffic back to a botnet controlled by the attacker. This evidence was instrumental in dismantling the botnet and prosecuting its operator.

4.1.3 Malware analysis

Reverse engineering of malware was another effective technique used to understand the functionality and intent of malicious software. A detailed examination of malware in a financial fraud case revealed a sophisticated keylogger that had been used to steal banking credentials. This analysis not only helped in attributing the attack but also informed the development of mitigation strategies.

4.1.4 Cryptocurrency tracking

The study highlighted the increasing importance of forensic tools designed for blockchain analysis. In a cryptocurrency theft case, investigators traced transactions across multiple wallets, ultimately identifying the perpetrators and recovering a portion of the stolen funds. Tools such as Chainalysis and CipherTrace were frequently cited as essential for such investigations.

4.2 Challenges in Digital Forensic Investigations

Despite the successes, the study uncovered several challenges that hinder the effectiveness of digital forensic investigations:

4.2.1 Encryption and data access

The widespread use of encryption presented a significant barrier to accessing digital evidence. Many cases required considerable time and resources to bypass encryption, delaying investigations and, in some instances, leaving critical evidence inaccessible.

4.2.2 Cloud-Based data

The shift toward cloud computing introduced complexities in data acquisition, particularly due to jurisdictional issues and the multi-tenant nature of cloud services. In one case, investigators faced difficulties obtaining evidence stored in a foreign-based cloud server, highlighting the need for international cooperation and standardized legal frameworks.

4.2.3 Diverse device ecosystems

The proliferation of IoT devices and diverse operating systems posed additional challenges. Forensic tools often required customization to handle unique device architectures and proprietary systems, increasing the technical demands on investigators.

4.2.4 Volume of digital evidence

The sheer volume of digital evidence in modern investigations created challenges in data processing and analysis. Several experts noted that existing forensic tools struggled to scale effectively, leading to delays and potential oversights in large-scale investigations.

4.3 Emerging Trends and Innovations

The study identified emerging trends that are shaping the future of digital forensic investigations:

4.3.1 Integration of artificial intelligence

AI and machine learning are increasingly being integrated into forensic tools to automate repetitive tasks, identify patterns in large datasets, and enhance decision-making. For example, AI-based anomaly detection systems have been deployed to flag suspicious activities in network logs, significantly reducing manual effort.

4.3.2 Focus on real-time forensics

The need for real-time forensic capabilities is becoming more apparent, especially in responding to active threats such as ransomware or insider breaches. Tools designed for live analysis are gaining traction, enabling investigators to collect and analyze evidence without disrupting ongoing operations.

4.3.3 Collaboration between stakeholders

Collaborative initiatives between law enforcement, private industry, and academia are fostering the development of more advanced forensic tools and methodologies. These partnerships are also facilitating knowledge sharing and standardization, addressing some of the challenges associated with jurisdictional and technological diversity.

4.4 Recommendations for Improvement

Based on the findings, several recommendations were proposed to enhance the effectiveness of digital forensic investigations:

4.4.1 Standardization of protocols

Developing and adopting standardized protocols across jurisdictions will improve consistency and facilitate international cooperation in cybercrime investigations.

4.4.2 Investment in advanced tools and training

Continuous investment in state-of-the-art forensic tools and specialized training for investigators is critical to keeping pace with evolving cyber threats.

4.4.3 Strengthening legal frameworks

Updating legal frameworks to address the challenges of encryption, cloud forensics, and cross-border investigations will enhance the ability of investigators to access and use digital evidence.

4.4.4 Scalable forensic solutions

The development of scalable forensic tools capable of handling large datasets will help address the growing volume of digital evidence in modern investigations.

5 DISCUSSION

The findings of this study provide valuable insights into the current state of digital forensic investigations in addressing cybercrime, revealing both the field's strengths and its challenges. This section discusses the implications of the results, explores their broader significance, and evaluates potential strategies to address identified gaps. The discussion also integrates perspectives from existing literature, expert insights, and case study findings to contextualize the role of digital forensics in combating cybercrime.

5.1 The Role and Effectiveness of Digital Forensics

Digital forensics has proven to be an indispensable tool in cybercrime investigations. The ability to uncover, analyze, and present digital evidence enables law enforcement and organizations to attribute attacks, recover stolen assets, and strengthen legal proceedings. For instance, the successful application of file system analysis and network traffic monitoring in the documented case studies demonstrates the field's capability to reconstruct complex cyber events and identify perpetrators. These findings align with prior research highlighting the importance of robust forensic methodologies in mitigating cyber threats.

However, the effectiveness of digital forensics depends heavily on the technical expertise of investigators and the quality of tools at their disposal. Advanced techniques such as malware reverse engineering and blockchain analysis require specialized skills and cutting-edge technologies. The increasing sophistication of cybercriminals underscores the need for ongoing investment in training and tool development to maintain the field's relevance and impact.

5.2 Challenges in Digital Forensic Investigations

The challenges identified in the study underscore the dynamic and evolving nature of digital forensics. Encryption, for example, poses a persistent barrier to accessing critical evidence, as cybercriminals continue to adopt advanced cryptographic techniques to secure their activities. While tools for encryption breaking exist, they often require significant time and computational resources, delaying investigations and, in some cases, rendering evidence inaccessible. This finding echoes concerns raised by [11] regarding the growing complexity of encryption in ransomware cases.

Cloud-based data introduces another layer of complexity. The distributed nature of cloud environments, coupled with jurisdictional differences, creates hurdles for evidence acquisition. Quick and Choo emphasized the need for specialized tools and legal frameworks to address these challenges, a sentiment echoed by the participants in this study. The increasing reliance on cloud services by individuals and organizations makes this issue particularly pressing[12].

The diversity of devices and platforms further complicates investigations. The rise of IoT devices and non-standardized systems often requires customized approaches to data extraction and analysis, stretching the capabilities of existing forensic tools. Additionally, the volume of digital evidence continues to grow exponentially, presenting scalability challenges that hinder efficient processing and analysis. These findings highlight the urgent need for scalable forensic solutions capable of managing large datasets without compromising accuracy.

5.3 Emerging Trends and Their Implications

The integration of artificial intelligence (AI) into forensic tools represents a promising trend. AI has the potential to enhance investigative efficiency by automating repetitive tasks and identifying patterns in complex datasets. For example, anomaly detection systems powered by machine learning algorithms can sift through vast volumes of network traffic to flag suspicious activities. However, the reliance on AI introduces new challenges, including algorithmic transparency and the potential for errors in automated analyses. Ensuring that AI-driven tools are rigorously tested and aligned with forensic standards will be critical for their successful implementation.

Real-time forensics is another emerging area with significant implications for cybersecurity. The ability to analyze digital evidence on-the-fly can provide valuable insights during active attacks, enabling investigators to respond more effectively. This shift toward proactive forensics requires not only technological advancements but also changes in investigative practices and workflows.

Collaboration between stakeholders—law enforcement, private industry, and academia—has been highlighted as a key enabler of progress in the field. Partnerships can facilitate knowledge sharing, drive innovation, and promote the development of standardized protocols. This collaborative approach is essential for addressing cross-border cybercrime, which often involves multiple jurisdictions and legal systems.

5.4 Recommendations and Their Feasibility

The recommendations proposed in this study are both practical and forward-looking. Standardization of protocols across jurisdictions is achievable through coordinated efforts by international organizations such as INTERPOL and the United Nations. These efforts should focus on creating uniform guidelines for evidence collection, handling, and analysis, ensuring consistency in forensic practices worldwide.

Investing in advanced tools and training is equally critical. Governments, organizations, and educational institutions must prioritize funding for the development of state-of-the-art forensic technologies and training programs. These investments will equip investigators with the skills and resources needed to tackle emerging cyber threats.

Strengthening legal frameworks to address the challenges of encryption and cross-border investigations is another vital step. Governments should work collaboratively to create agreements that facilitate evidence sharing and streamline the legal processes involved in cybercrime cases. Legislative updates must also account for the unique characteristics of digital evidence to ensure its admissibility in court.

Finally, scalable forensic solutions must be prioritized to address the growing volume of digital evidence. Cloud-based forensic platforms and distributed computing systems can provide the necessary scalability while maintaining accuracy and efficiency. Collaborative research initiatives can play a significant role in advancing these technologies.

6 CONCLUSION

The discussion highlights the critical role of digital forensics in addressing cybercrime while emphasizing the need for continuous innovation and adaptation to overcome emerging challenges. By addressing gaps in tools, training, and legal frameworks, digital forensics can maintain its effectiveness in an increasingly complex digital landscape. The integration of new technologies and the fostering of collaborative relationships will be instrumental in ensuring that forensic investigations remain a robust and reliable response to cyber threats.

Digital forensic investigation is an indispensable tool in combating the growing menace of cybercrime. By enabling investigators to collect, analyze, and present digital evidence, it plays a crucial role in identifying perpetrators, uncovering attack methodologies, and supporting legal proceedings. This study examined the application of digital forensic techniques in cybercrime cases, explored the challenges faced by investigators, and proposed actionable recommendations to enhance the field's effectiveness.

The findings revealed that digital forensic techniques, including file system analysis, network traffic monitoring, malware reverse engineering, and cryptocurrency tracking, are highly effective in reconstructing events and identifying key evidence. However, significant challenges persist, such as encryption barriers, the complexities of cloud-based environments, and the scalability demands of processing large volumes of digital evidence. Jurisdictional hurdles further complicate investigations, especially in cases of cross-border cybercrime.

Emerging trends such as the integration of artificial intelligence, the focus on real-time forensics, and enhanced stakeholder collaboration present promising opportunities to address these challenges. AI-powered tools can improve efficiency and accuracy, while partnerships among law enforcement, private industry, and academia can drive innovation and standardization.

To strengthen the field of digital forensics, the study recommends several strategies, including the standardization of investigative protocols, increased investment in advanced tools and training, the development of scalable forensic solutions, and the establishment of robust international legal frameworks. Implementing these recommendations will enable digital forensics to remain a critical component in the fight against cybercrime.

In conclusion, as cyber threats continue to evolve in complexity and scale, the importance of digital forensic investigations cannot be overstated. Through innovation, collaboration, and strategic investment, digital forensics can adapt to emerging challenges and uphold its role as a cornerstone of modern cybersecurity and justice systems.

CONFLICT OF INTEREST

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Sihombing, E, Erlina, Rujiman. The effect of forensic accounting, training, experience, work load and professional skeptic on auditors ability to detect of fraud. International Journal of Scientific and Technology Research, 2019, 8(8): 474-480.

[2] Abdullah, I, Lubis, A W, Sumitra, A. Explanation of Forensic Accounting and Its Application (Case Some Industry Sector). Journal of Pharmaceutical Negative Results, 2022, 1585-1588. DOI: https://doi.org/10.47750/pnr.2022.13.S09.195.

[3] Khamidovich, K B, Zokirovich, K B, Mirshokhidovna, T D. General Theoretical Issues of Improving Private Forensic Methods In The Field Of Combat Against Cybercrime. Psychology and education, 2021, 58(1): 2705-2712.

[4] Yaacoub, J P A, Noura, H N, Salman, O, et al. Digital forensics vs. Anti-digital forensics: Techniques, limitations and recommendations. 2021. arXiv preprint arXiv:2103.17028.

[5] Choi, K S, Back, S, Toro-Allvarez, M M. Digital forensics and cyber investigation. Cognella. 2023.

[6] Hemdan, E E D, Manjaiah, D H. An efficient digital forensic model for cybercrimes investigation in cloud computing. Multimedia Tools and Applications, 2021, 80, 14255-14282.

[7] Yaacoub, J P A, Noura, H N, Salman, O, et al. Advanced digital forensics and anti-digital forensics for IoT systems: Techniques, limitations and recommendations. Internet of Things, 2022, 19, 100544.

[8] Casino, F, Dasaklis, T K, Spathoulas, G P, et al. Research trends, challenges, and emerging topics in digital forensics: A review of reviews. IEEE Access, 2022, 10, 25464-25493.

[9] Tok, Y C, Chattopadhyay, S. Identifying threats, cybercrime and digital forensic opportunities in Smart City Infrastructure via threat modeling. Forensic Science International: Digital Investigation, 2023, 45, 301540.

[10] Yeboah-Ofori, A, Brown, A D. Digital forensics investigation jurisprudence: issues of admissibility of digital evidence. Journal of Forensic, Legal & Investigative Sciences, 2020, 6(1): 1-8.

[11] Alazab, M, Layton, R, Broadhurst, R, et al. Malicious spam emails developments and authorship attribution. In 2013 fourth cybercrime and trustworthy computing workshop . IEEE. 2012, 58-68.

[12] Quick, D, Choo, K K R. Big Digital Forensic Data: Volume 2: Quick Analysis for Evidence and Intelligence. Springer Singapore. 2018.

Multidisciplinary Journal of Engineering and Technology
ISSN: 3007-6900
DOI: https://doi.org/10.61784/mjet3018

INTEGRATING INDIGENOUS KNOWLEDGE SYSTEMS INTO ADVANCED MATHEMATICAL FORMULA DEVELOPMENT: A FRAMEWORK FOR CURRICULUM INNOVATION IN SOUTHERN AFRICA

Stephen Kelvin Sata
ICOF Global University, Lusaka, Zambia.
Corresponding Email: stephensata@gmail.com

Abstract: This doctoral research takes a step further to explore how IKS could be deployed to enhance the creation of even more complex formulas for advanced mathematics curricula, thereby providing a sound framework for curriculum enhancement in Southern Africa. It aims at bridging gaps of what Indigenous mathematics has yet to be allowed to contribute to modern mathematics education and higher education development. We use case studies, ethnomathematics, post-colonial education theory, and curriculum development models as discourses in this research to explore how IKS may potentially recast the rephoneticised formulation of mathematical equations in line with the socio-cultural realities of southern Africa.

For this reason, the proposed framework revolves around three key elements. First, it supports conceptualizing and cataloguing Indigenous knowledge in mathematics and their rough appreciation as knowledge systems emanating from Indigenous peoples. Second, it focuses on integrating these ideas into current mathematical contexts—within the current storyline of mathematics, integration between traditional and innovative mathematical approaches. Last, it captures the re-contextualization of educational content in collaboration with Indigenous people, scholars, educators, and policymakers for relevance, equity, and practicality.

This work also explores the issues involved in implementing IKS in formal education, where resources are scarce; Indigenous knowledge and skills are not valued; culture and timely entry are issues that are not given adequate attention, and there is a critical need to prepare teachers through other professional development programmes to enable them to provide for inclusive education. It outlines an approach to addressing these concerns, including creating culturally appropriate knowledge-sharing Radicals, co-designing curriculum with communities, and pursuing policies of equity.

The research states that integrating IKS in formula derivation can redefine education through increased students' participation, enhanced diversity learning, and ultimately advocating Indigenous cultures. Moreover, it situates this integration as an enabler of future-oriented Implications, a new way of thinking mathematically for solving global problems using the architecture of both old and new knowledge systems.

In conclusion, this work aims to enable the sharing of knowledge between indigenous and formal education systems towards prescribing radical changes to curriculum in Southern Africa. It imagines a future in which mathematical education not only represents but also appreciates various forms of knowledge in the region aenhancing the learning climates suitable for the global community. This work forms a platform on which postgraduate scholars, educators and policymakers can collectively reimagine and reposition Indigenous knowledge within advanced mathematical learning and academic pursuit.

Keywords: IKSM; Continuing the work of formulating and applying mathematical and other useful formulas; curriculum development; Southern Africa

1 INTRODUCTION

Mathematics is often perceived as a universal discipline, yet its development and application have been deeply influenced by cultural, historical, and regional contexts. In Southern Africa, as in many parts of the world, indigenous knowledge systems (IKS) have historically played a significant role in shaping mathematical thought and practices. However, colonial educational systems often marginalized these systems, leading to a lack of representation in formal curricula [1]. This study aims to bridge the gap between traditional knowledge and contemporary mathematics by exploring how IKS can inform the development of advanced mathematical formulas and contribute to culturally inclusive curriculum innovation in Southern Africa.

Indigenous knowledge systems encompass a wide range of local practices, techniques, and wisdom accumulated over generations [2]. In the mathematical domain, these systems include indigenous counting methods, geometric patterns, and spatial reasoning, which have historically been applied in activities such as architecture, navigation, and agriculture [3]. Despite their potential value, these contributions are often excluded from mainstream education, which predominantly adopts Western-centric mathematical frameworks [4]. This exclusion not only limits the richness of mathematical

© By the Author(s) 2025, under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0).

knowledge but also alienates students who might find greater resonance with mathematical concepts rooted in their own cultural contexts [5]

Mathematics is often seen as originating from the basic needs of ancient societies, such as counting, measuring land, and tracking time. The earliest evidence of mathematical activity dates back to around 3000 BCE in ancient Mesopotamia, where the Sumerians and Babylonians developed systems of arithmetic and geometry. According to Burton (2011), early mathematics was primarily concerned with practical tasks like accounting, agricultural measurement, and astronomy.

Mathematics in Ancient Egypt and Greece: In Ancient Egypt, the practice of mathematics was also applied to tasks such as building pyramids and managing resources. The Egyptians were among the first to use a system of geometry for land surveying [6].The Greeks, particularly through figures like Euclid, Pythagoras, and Archimedes, transformed mathematics from a practical tool to an abstract discipline. Euclid's "Elements" (circa 300 BCE), a foundational text in geometry, laid the groundwork for deductive reasoning in mathematics. The Greek approach to mathematics was marked by a systematic organization of mathematical knowledge through axioms and proofs [7].

Islamic Golden Age Contributions: The Islamic Golden Age (8th–14th centuries) played a pivotal role in preserving and expanding mathematical knowledge. Scholars like Al-Khwarizmi made major contributions to algebra, and his book "Al-Kitab al-Mukhtasar fi Hisab al-Jabr wal-Muqabala" introduced systematic methods for solving linear and quadratic equations. This work laid the foundations for modern algebra and influenced European mathematics during the Renaissance [8].

Mathematics in the Modern Era: The development of modern mathematics took a significant leap with the advent of calculus, independently discovered by Isaac Newton and Gottfried Wilhelm Leibniz in the late 17th century. Their work is often credited with shaping the course of modern mathematics. The invention of calculus marked a pivotal moment in the history of mathematics, as it allowed for the study of continuous change, setting the stage for later developments in physics, engineering, and economics [9].

Recent advancements in ethnomathematics, the study of the relationship between mathematics and culture, provide a foundation for integrating IKS into formal education. Highlights the importance of contextualizing mathematical education to reflect local knowledge systems, arguing that such approaches enhance student engagement and foster critical thinking. This study builds on these principles by proposing a framework for incorporating IKS into the design of advanced mathematical formulas, with a focus on curriculum innovation.

The framework addresses three key areas: documenting indigenous mathematical concepts, aligning these concepts with existing mathematical frameworks, and developing collaborative educational materials involving indigenous communities, educators, and policymakers. By situating mathematical education within local contexts, this approach not only enriches the discipline but also serves broader goals of decolonization and sustainable development [10].

In this study, we examine the potential of IKS to inspire innovative approaches to mathematical formula development in Southern Africa, addressing challenges such as cultural sensitivity, resource constraints, and capacity building for educators. Through this exploration, we aim to contribute to the ongoing dialogue on decolonizing education and creating curricula that are both globally competitive and locally relevant.

2 LITERATURE REVIEW

Incorporating Indigenous Knowledge Systems(IKS) into curricula of modern education systems has also come into focus, especially when education systems are looking for suitable ways of diversifying their systems to de-colonies the limited Westernized global informative systems. Due to diverse traditional practices and beliefs among various groups of indigenous people of Southern Africa, a lot of potential exists as far as provable indigenous knowledge leading to the formulation of better mathematical equations is concerned. This integration can enhance the curriculum, make mathematics more relevant, and be close to the students and society. This literature review examines the theoretical practices of integrating IKS into advancing mathematics curricula in Southern African countries.

2.1 Indigenous Knowledge Systems in Mathematics

IKS is a specialized form of cultural heritage knowledge, discipline, and wisdom encompassing a community's practices and beliefs, often conventional, repetitive, and traditional. These systems entail a worldview that integrates knowledge with physical and cultural context and responds to the needs of the local context [2]. IKS, therefore, is not only about who and how it provides solutions to peoples' immediate problems but also about a whole paradigm with epistemology and ontology for understanding and interacting with the physical environment. In mathematics, indigenous knowledge systems harbour considerable resources that can enrich and complement formal education on geometry, measurement, patterns and space relationships [9].

From the same sector of mathematics, the indigenous knowledge of mathematics is always expressed in the paradigms of daily experiences, and it possesses and captures the mathematical knowledge in natural form, which is still a confirmation of the professional mathematical thoughts that are contrary to the total formal education and artificialities that belongs to the western set of knowledge. In his opinion [11], continues that these Indigenous methods, as conveyed by Indigenous communities, are not mere fables or real-life examples but systematic approaches Indigenous communities use to reason mathematically. For example, counting methods that have been in use in different African communities, for instance, the Shona of Zimbabwe or the Zulu of South Africa, are complex and far from the methods of linear counting but include other aspects such as grouping and modular arithmetic [9]. These counting systems embody number and operation schemata that can extend what students learn formally in mathematics classrooms by presenting other ways of counting.

Besides, counting systems comprise Indigenous mathematical knowledge: geometric architecture, land measurement and Indigenous navigation. For instance, elements such as geometric design and orientation related to African construction are critical in establishing appropriate and attractive structures and have always been the focus of traditional African architecture. Which, captures how geometric shapes have been used in the construction of huts and ceremony areas in Southern Africa; there was defined knowledge of symmetry and proportion in the construction of structures apart from aesthetics of shapes [11]. These architectural practices illustrate how geometry can be taught using real-life practices, thus facilitating students' teaching of geometrical concepts such as symmetry, area, and volume.

Also, astronomical knowledge embedded in the indigenous cycles provides mathematically rich content. Astronomical phenomena have been known for centuries by many indigenous peoples to plan their planting, navigation and calendars. For example, the indigenous people of Southern Africa, San, have learned to use the position of the stars and seasons to calibrate their agricultural activities [7]. Comprehension of phases of the moon and positions of stars was applied in practical endeavors and in ways of the timely division of social activities. That kind of knowledge has a realistic foundation that can be used to demonstrate complicated topics such as trigonometry, angles, and periodic functions.

Incorporating such indigenous practices into the formal school system offers a good opportunity to supplement the arguably superior compilation of advanced mathematical student knowledge. Using IKS in the curriculum can always come with different approaches to solving mathematic issues and enhancing learners' creation and critical minds [12]. As much as it helps the students identify with their own cultural knowledge systems that are harnessed when teaching mathematics, they get motivated to learn the content. In addition, this integration enhances Students' understanding of Mathematics because they can apply the knowledge they are learning in the classroom to their reality/point of view.

Integrating IKS into Mathematics syllabi can also assist in closing the gap between Mathematics and students' reality. For instance, counting patterns familiar to Indigenous people can replace number theory, or explaining symmetry with African art geometric design can make it more real to students. This approach also improves students' knowledge of mathematics and values Indigenous people's knowledge as an important and diverse part of intellectual production.

This is particularly the case in Southern Africa, where colonial and post-colonial education have influenced education; therefore, using IKS to teach mathematics is also a mechanism of decolonizing education. Most academics assert that education decolonization is the process of recapturing education by questioning how knowledge is constructed, disseminated, and accredited. When incorporating indigenous knowledge into education, the teacher not only undermines the monopoly of the Western systems of mathematics but also helps students see that we can indeed count on the indigenous way. This shift helps indigenous students and enhances the understanding of mathematics education as a multicultural topic [13].

Therefore, more research concerning including IK in advanced mathematics curricula in Southern African countries remains a rich potential for creating meaningful mathematical education. Even counting systems, geometry, and astronomical knowledge, which are part of Indigenous practices, can be implemented into the learning-teaching process to offer students a much richer and culturally diverse view of mathematics. Besides promoting growth in students' math, this approach supports the growth of critical thinking, innovation, and cultural engagement, thus enabling a more non-sexist and de-imperialized education system. Indigenous knowledge systems in mathematics can been seen in Figure 1

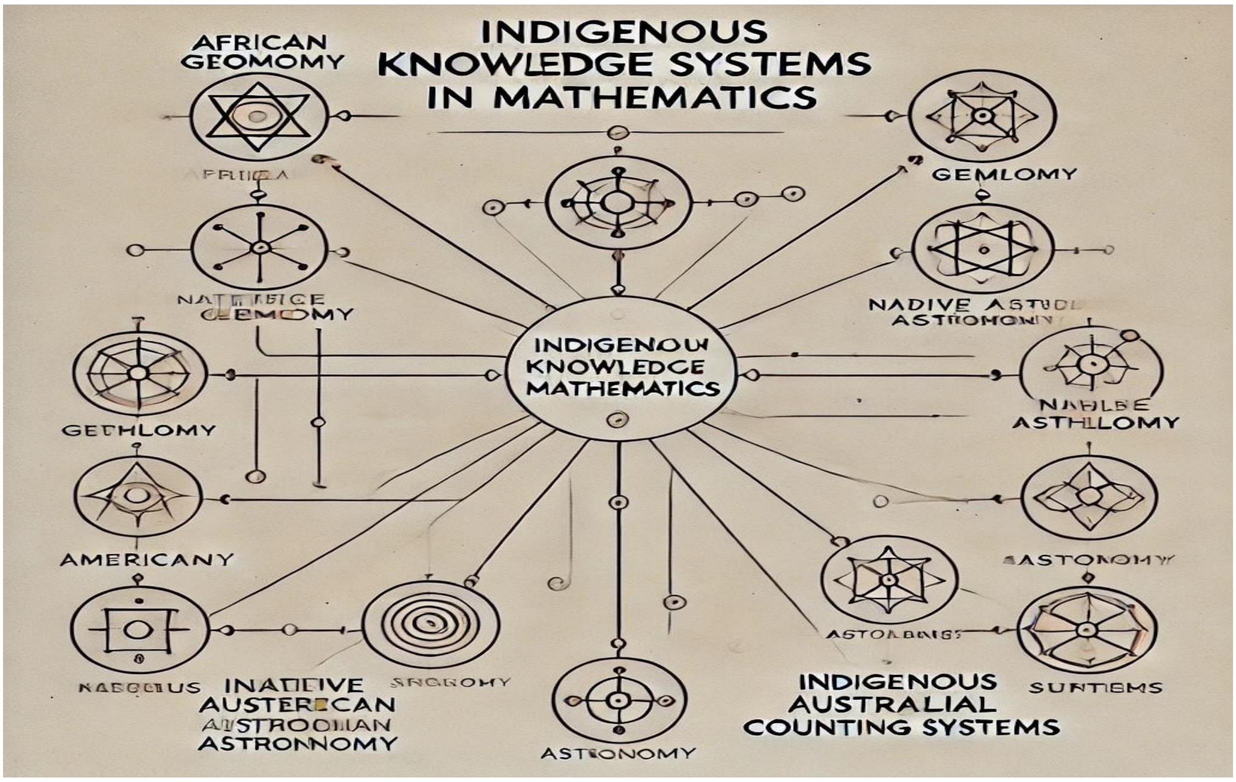


Figure 1 Indigenous Knowledge Systems in Mathematics

3 CURRICULUM INNOVATION IN SOUTHERN AFRICA

Curriculum development in Southern Africa has indeed turned into a subject of interest in teaching reforms in the past few years, where more emphasis is placed on the reformulation of the historical impacts of colonialism and apartheid. The history of such socio-political systems produced imprints on education, especially in how curricula were developed and practiced. The call for curriculum development, especially after the decolonization process, aims to provide a watermark to the traditional imperialism-oriented Western perspectives on education. Consequently, scholars and educators in the region have called for adopting and implementing IKS in school curricula to extend education for everybody, ethnocentric and multicultural, so as to be sensitive to the diversity of learners.

Integrating IKS into the curriculum is part of a more extensive DE colonial course that seeks to unite educational systems with communities of color and their cultural and social histories. , Such curriculum innovation produces one of its primary objectives of narrowing the gap between the knowledge taught in school and young people's knowledge and experiences [1]. IKS enables the curriculum developer to design meaningful instructional material for students as it blends with their daily experiences. The link between formal education and indigenous knowledge can improve how students embrace their cultural value system while enabling them to adapt to the prevailing world.

In the advanced level of mathematics curriculum, when learning becomes more concrete and specialized, incorporating IKS provides an enriched curriculum development area. Thus, mathematics education, frequently viewed as a formal and timeless science, could be enriched with native numerical and geometrical knowledge and skills, theories, and problem-solving skills developed over generations among people of various African tribes. This approach also reacts against mathematics as a prerogative of, let alone as a creation of, the Western world only. It establishes the view of mathematics as something constructed within cultural and historical settings. In this way, the students can increase their knowledge of mathematics and thereby obtain a global conception integrated with the local worldview.

It has also been notice that there is a need for curriculum reforms that integrate modem mathematics with African mathematical knowledge [14]. They claim that a good curriculum can indicate the development of mathematical theory on the international level and consider the local mathematical practices and knowing practices that were muted in mainstream education. For example, there are mathematical and problem-solving practices within some African contexts, and mathematical and problem-solving in the African context have long been cultural communities' practices in their social practices. This is why, by integrating such knowledge into the curriculum, the student can learn how to consider mathematics as a fluid cultural phenomenon of knowledge.

Nevertheless, the introduction of IKS into curriculum development in Southern Africa has its drawbacks. However, there is a significant hindrance regarding the absence of best practices for addressing the integration of Indigenous Knowledge

Systems in curricula. Nevertheless, there is a general agreement about the significance of IKS; however, there needs to be more agreement on how it could be implemented systematically in school education. However, most teachers in the Southern African context need more training on mainstreaming IKS in their teaching practices across all academic fields, especially in more complicated areas such as mathematics. More instructional training must be needed to ensure the achievement of curriculum reforms at the teacher level. , as other scholars argue [13], when not well trained, teachers may have a hard time incorporating IKS in ways that are effective in helping the students learn and feasible.

The other challenge is in educational institutions where there is still much reluctance to change most classes, which still need to be centered on the West. This resistance is less about the implementation process than with members of educational policy-making, who may be highly reluctant to bring about profound changes to the system. Such resistance arises from structural factors within the education systems that are still discriminatory in promoting Western knowledge over indigenous knowledge systems [14]. Addressing these challenges, therefore, means going beyond organization and policy support and, in principle, making society receptive to valuing Indigenous knowledge.

Nevertheless, applying IKS in education systems is the right way to enhance culturally relevant philosophy in education in Southern Africa. In fact, by embracing mathematics as a system of knowing the world that has developed in various cultures, curriculum developers can deconstruct the narrow and oppressive way mathematics is taught and learnt. In other words, what is envisaged is offering students an education in mathematics based on al-intimacy al-Adam al-calamity in a way that affirms and values the human endowment for cultural and intellectual capital in the region. This approach presents the probability of the progressive nature of an educational system that equally benefits and addresses the needs and goals of all learners.

Therefore, embedding IKS into the advanced mathematical curriculum in Southern Africa presents a suitable strategy towards removing colonialism and the relevance of education to the sociocultural realities of learners. As with any new model, some concerns should be considered; however, there are also advantages to such curriculum development, including increasing the students' appreciation for local cultures, increasing their interest and encouraging the creation of new, more culturally sensitive pedagogy. This points to the objective homework of enabling South and Southern Africa for ergonomics of education that integrates advanced divergent science and indigenous knowledge for higher learning for all students [1] Curriculum Innovation in Southern Africa can be seen in Figure 2.

.

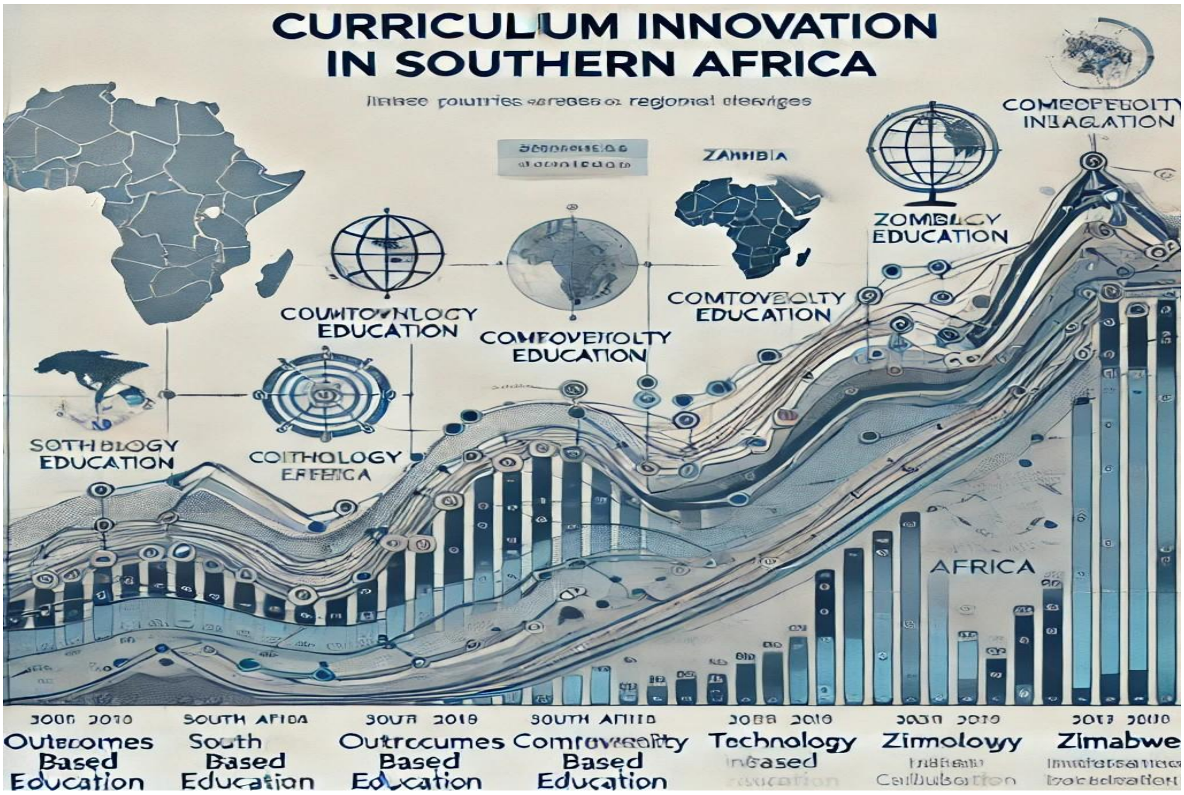


Figure 2 Curriculum Innovation in Southern Africa

3.1 Indigenous Knowledge Systems and Early Mathematical Practices

Curriculum development in Southern Africa has indeed become an area of interest within the teaching reforms in the past few years, emphasizing the historical legacies of colonialism and apartheid. The impacts of the various socio-political systems in their individual histories left some marks on education and curriculum development and practice. The cry for curriculum development and, more so, after centering on delo-colonialization, seeks to bring an antidote to the more traditional imperialism-inclined views on education within the Western world. Subsequently, scholars and educators in the region have recommended that IKS be incorporated into school curricula to make education for everybody ethnocentric, multicultural, and responsive to the learners' diversity.

Therefore, implementing IKS into the literacy curriculum is a part of an even more significant de colonial coursework that aims to link educative systems to people of color and their cultural and historical roots. Curriculum innovation achieves one of the central goals of countering the disjuncture between curricular knowledge and young people's knowledge and experiences [14]. IKS makes it easier for curriculum developers to develop meaningful lesson content for students since it integrates with their daily activities. The combination of formal education in school with indigenous knowledge will enhance how the students appreciate and embrace their Afrocentric values- referent to self and social reality while preparing them for the current global world.

When the learning becomes concrete and unique at the advanced level of mathematics curriculum implementation, the incorporation of IKS makes an enhanced curriculum development area. Then, mathematics education, which could be formally and academically defined as an ancient and universal science, benefitted from native numerical and geometrical knowledge and skills, theories, and problem-solving skills nurtured in different generations of African tribes. It also responds to mathematics as a privilege of, or better still, a product of, the Western world alone. It creates the perspective of mathematics incarnate within cultural and historical frameworks. Thus, the students can deepen their knowledge of mathematics and, thus, receive the global conception joined with the local view.

It has also been observed a need to transform the curriculum to embrace modern mathematics within African Mathematics knowledge. They argue that they understand that a good curriculum can point to the satisfaction of developing mathematical theory on the international level, responding to the local mathematical practices and knowing practices that were erased in mainstream education [15]. For example, there are mathematical and problem-solving practices in some African cultures, and mathematical and problem-solving practices in the African cultural community have always embraced social practices. This is why, by incorporating such knowledge into the programme, the student can develop expertise in approaching mathematics as a cultural phenomenon of knowledge in learning.

However, as with any new idea, introducing IKS into curriculum development in Southern Africa has its demerits. However, there is a significant challenge to the need for existing guidelines to address incorporating Indigenous Knowledge Systems into curricula. However, there is much consensus about the importance of IKS, and much more is needed on how it should be introduced systematically into school education. Nonetheless, most teachers in Southern Africa require further professional development in mainstreaming IKS in their teaching across all disciplines, particularly in the more complex disciplines of content areas such as mathematics. Probably more instructional training might be required to address the achievement of curriculum reforms at the teacher level, as other scholars pointed out that if not adequately trained, teachers may struggle to incorporate IKS in a manner that assists the students to learn and can easily be implemented [15].

The other is that there is still so much resistance in educational institutions, especially in changing most classes, which must still be oriented towards the West. This resistance is not as linked to the implementation process as some educational policy-making entities and may be averse to changing the system. Such resistance stems from such structures within the education systems that are discriminatory in presenting Western knowledge as superior to indigenous knowledge systems [14]. Mitigating such challenges requires moving from the organization and policy support level and, in principle, sensitizing society to appreciate Indigenous knowledge.

However, the practice of IKS in education systems is geared towards increasing culturally appropriate philosophy in the education system in Southern Africa. Indeed, by identifying mathematics as a form of knowing the world that exists in different cultures, curriculum developers can disestablish the suppressed and oppressive positivist view of mathematics being taught and learnt in most education systems. In other words, what is envisaged is the provision of students with an education in mathematics in the light of al-intimacy al-Adam al-calamity in a positive and culturally valuable way with an appreciation of the human endowment for cultural and intellectual capital in the region. As such, this approach exposes the possibility of an educational system's progressive features that provide equal value and services to accomplish learners' goals and needs.

Thus, the integration of IKS in the advanced mathematical curriculum in Southern Africa is the right approach towards the elimination of colonialism and practices relevance of education in the learner's sociocultural context. Like any new strategies, specific issues should be taken into account; nevertheless, the blessings of such curriculum development are the following: the students are offered more opportunities to value a culture of the particular region, and such value can stimulate and grow the students' interest to develop new, more sensitive to local context, pedagogy. This shows the research needed for opening homework for South and Southern Africa for ergonomics of education that restructures the decolonized higher learning for the students[14-15].

3.2 Introduction of Western Mathematics through European Colonization

The introduction of Europeans to Southern Africa during the seventeenth and eighteenth centuries signified the beginning of a shift in the educational structure of that region, especially in Mathematics. European powers, mainly the British and the Dutch, extended institutional education in Europe and brought about a revolution in the architecture of education in the colonial context. During this period, the European systems of education and math were adopted, and practices were totally devoid of the black-induced knowledge that had been present before colonization.

In Southern Africa, the initial style of colonial education was mainly set to prepare a few selected locals fit to serve colonial masters' purpose. Essentially, this education entailed skills that would cultivate trade capacities that would underpin the colonial economy, especially arithmetic, bookkeeping, and other forms of field mathematics connected to commerce, trade and resource mobilization. Much credit must be given to the missionary agencies which accompanied missions in the region, especially in the 19th century, to set up schools that taught rudimentary mathematics of the European type. Teachers imposed known European applications of mathematics and devised curricula that would prepare colonials for administrative careers as the Continent's rulers rather than instilling in them general mathematical education.

In the past, has been highlighted that colonial mathematics education was designed to best suit the economic and administrative needs of the colonizing nations rather than for the childhood development of the colonized indigenous people [14]. It is also important to recognize that this education system disadvantaged the remaining and far more numerous more significant Indigenous peoples who had similarly restricted education and limited access to more advanced mathematical learning. The system ensured only mere arithmetic skills and bookkeeping; thus, as it applied, only survived; a minority of children of settlers and local elites were privileged enough to receive formal arithmetic lessons. The rest of the indigenous people, on the other hand, had, in most cases, never any chance to even learn any formal education, let alone mathematical concepts.

Moreover, the concentration of European mathematics in commerce strained the connection between European mathematics and the original math techniques that existed before colonization in the African Continent. Pre-colonial Indigenous Africans had quantitative knowledge that was based on the conventional modes of handling their day-to-day experiences: field planting, animal rearing, and communal ceremonies. However, these systems were not embraced in the colonial education regime, and this contributed to the denial of Indigenous knowledge's intelligence compared to European knowledge.

The current impact was a split between conventional learning and traditional African knowledge. According to Mwaura (2019), the non-recognition of Indigenous mathematical traditions as part of the teaching-learning curriculum kept African cultures and knowledge systems in disarray. A generation of learners who were trained in European mathematics emerged but had no connection to the mathematical practices within their communities.

Therefore, the establishment of a mathematical curriculum in Southern African countries was due to European colonial education systems imported in the 17th and 18th centuries, which aided in the provision of education only for colonial needs. This system provided an enhanced thorough mathematical computation for mercantile industries, but a strand of higher education was barred for the Indigenous community. The above outcome manifested in a disconnection between formal mathematics education and indigenous knowledge systems that persist in shaping education systems in many African nations.

4 POST-COLONIAL MATHEMATICS EDUCATION AND INTEGRATION OF INDIGENOUS KNOWLEDGE

In the middle of the twentieth century, most independent Southern African countries had to conduct considerable policymaking assessments of the educational systems they inherited from the colonial powers and start over to reconstruct their theories and values. Speaking at the symposium, scholars and educators highlighted some native knowledge systems to be adopted when developing educational systems relevant to the region and hence the call to adopt Indigenous knowledge systems, especially when developing education structure and content, as seen when the education theorists suggested ways through which integrated education structure could be promoted especially when adopting Indigenous knowledge systems in the course curriculum. This movement was geared towards redressing the education colonial legacies that had excluded Indigenous/post-colonial societies and equally aimed at transforming post-colonial societies by recognizing Indigenous knowledge/intelligentsia. The assimilation of Indigenous knowledge into mainstream education was therefore presented as the right move in the effort to de-exploit education, especially with relation to curricula that have their origin in the Western world.

Indigenous knowledge systems became one of the central features of the post-colonial curriculum reform within mathematics education. Van der Walt has postulated that, on the one hand, many Southern African countries remained anchored in Western mathematics through educational systems; on the other hand, the importance of Indigenous knowledge for mathematics education was gradually unveiled. This recognition was not viewed symbolically but as making mathematics more culturally responsive to students from different backgrounds. Consequently, as Van der Walt (2018) defines it, Indigenous mathematical knowledge encompasses more than an accumulation of specific mathematical practices and concepts revealed in the Indigenous communities and entwined in their culture and language practices. Within this context, mathematics was approached not as the ahistorical edifice people often conceive it to be but as a field with its origins, resources, approaches, and learning experiences as situated in context.

In South Africa, for example, attempts to introduce Western and Indigenous mathematical practices to students that were well integrated were made under the project of the decolonizing curriculum. Such attempts were an attempt to close the gap between formal logical European mathematics and contextually situated mathematics of cultures indigenous to Africa . Even as indigenous knowledge was positioned as a key priority for constructing an equitable mathematics curriculum, it was recognized that incorporating it would profoundly transform teaching and learning practices. This vision was to recognize the Indigenous knowledge system as an alternative to the Western mathematical system. However, the approach should complement the two so that other students who do not relate to Indigenous mathematics can benefit from the knowledge.

Chikoko et al. opine that convictions to integrate Indigenous mathematical practices into formal curricula have faced challenges. A lack of professional development among the primary facilitators for teachers to competently address the Western and Indigenous knowledge systems is pinpointed. Lack of resources, professional development, and skills among teachers, especially in rural and impoverished schools, means that these researchers' frameworks and frameworks remain out of reach for many teachers. There is more of this deficiency because the broader education system, in terms of its framework, still draws from the Western education model and has little space for indigenous knowledge systems even now . More so, as Chikoko et al. have pointed out, policymakers are most likely to resist change with regard to such incorporation since they may regard indigenous knowledge as inferior or incompatible with the standards of foreign institutions. This resistance comes from a firmly held perception of the Western form of education as the yardstick for education and development, informing educational restructuring processes in many Southern African nations.

Still, continuous discussions of decolonizing mathematics instruction demonstrate the possibility of introducing a new perspective devoid of ethnocentric assumptions and better aligned with the local culture. The effort to include indigenous knowledge in the mathematics classroom is not only a scholarly process; it is a political and social act since it attempts to correct the harms done by colonial education systems. Despite these hurdles, educationists, policymakers and scholars have remained pallets for curriculum changes that incorporate the mathematical cultures of Southern Africa to develop an academic-oriented curriculum that at the same time respects culture.

4.1 Mathematics Education in the Contemporary Context

However, in the past decade or so, particularly in Southern Africa, there has been growing emphasis on using indigenous knowledge to deconstruct colonialism in education. The author, such as Maphosa, supports the rationale that including indigenous mathematical knowledge can improve the likelihood of mathematics for students from cultural backgrounds. Indigenous practices, counting methods, geo, metrics and patterns found in African art, to name but a few, can be incorporated into the curriculum and make the content area more culturally relevant for the students.

According to Nhemachena and Jansen, there is a tendency to adopt indigenous knowledge practices alongside the modern approach to theoretical mathematical knowledge in an educational setting. For instance, the conventional African forms of geometrical shapes or the initial abacus of counting patterns used by blacks can form the foundation for teaching such sets as algebra or fractals. This approach not only recognises the mathematical achievements of indigenous cultures but also contributes to creating a better social environment that accepts students as members of their cultural groups.

4.2 Benefits of Integrating IKS into Advanced Mathematical Education

Incorporating IKS in advanced mathematical education has several educational and cultural values. A chief gain is that mathematics becomes more familiar and understandable for learners from indigenous peoples because methods of pedagogics to teach math seem irrelevant to indigenous learners. Due to the utilization of IKS, students can establish mathematics links about their cultures, as postulated by Van der Walt. The indigenous patterns of design and the traditional system of measurement can be taken as the basic framework on which the learners may find the concepts of symmetry, algebra and calculus more interesting.

Moreover, there will be an improved understanding of the inclusion of indigenous philosophies into mathematics embraced by students so that they will see the cultural value and global aspects of mathematics. Besides extending students' mathematical knowledge, this two-pronged strategy enhances the children's self-identity and cultural associations, as Maphosa has pointed out. The social validity of Indigenous knowledge in mathematics as a tool for improvement may help empower and enhance the student's motivation to enhance their education through post-graduate levels in Mathematics and related fields..

4.3 Challenges in Integrating IKS into the Mathematics Curriculum

Nevertheless, including IKS in advanced mathematics education in Southern Africa is challenging. One challenge is that there needs to be a systematic and coherent pattern for integrating Indigenous knowledge into modern mathematics curricula. Most education systems continue only to acknowledge and encourage Western ¡ñ mathematical culture, thus hardly providing room for local knowledge. The authors support this by saying that Chikoko et al. argue that educational authorities and policymakers always resist Indigenous knowledge since they consider it inferior and less scientific than Western knowledge.

This is relatively rare when compared to the other challenges that have been discussed above. There is also a need for more professional development in teachers. Teachers may need more competencies to make indigenous knowledge systems acceptable to supplement formal mathematical learning. This implies that there is a need for proper training of teachers through training colleges to have the appropriate knowledge and methods regarding indigenous knowledge and modern views on mathematics. Furthermore, there are still limitations to the availability of resources such as text and other teaching materials, including texts that fill the interface between IKS and higher-level mathematics.

4.4 Opportunities for Curriculum Reform and Teacher Development

Despite these challenges, significant opportunities exist for curriculum reform and teacher development. There is growing recognition of the value of Indigenous knowledge in educational contexts and an increasing willingness to explore alternative pedagogies. Nhemachena and Jansen suggest that partnerships between Indigenous knowledge holders, mathematicians, and curriculum developers could facilitate the creation of a curriculum that balances Indigenous and formal mathematical knowledge.

Professional development programs for teachers are another important opportunity. These programs should equip educators with the knowledge and tools to integrate IKS into their teaching practice. Collaborative workshops and the development of resources that combine indigenous knowledge with advanced mathematical principles can also help educators gain confidence in delivering such a curriculum.

Integrating Indigenous Knowledge Systems (IKS) into teaching practice requires robust teacher training that focuses on specific content and comprehensive resource support. Training programs should include modules on understanding indigenous knowledge, covering its origins, principles, and applications in subjects like mathematics, science, and environmental studies. Teachers should be equipped with culturally responsive pedagogical strategies that enable them to contextualize lessons using storytelling, metaphors, and local cultural examples. Additionally, workshops on curriculum co-design should guide educators in developing lesson plans that blend IKS with modern academic content, with input from indigenous knowledge holders to ensure authenticity. Training should also address culturally sensitive assessment techniques to evaluate students' understanding of both indigenous and conventional knowledge.

Resource support is equally essential for successful integration. Teachers need access to culturally relevant teaching materials, such as textbooks, multimedia resources, and activity guides tailored to specific regions. Platforms for collaboration with indigenous elders and cultural practitioners should be established to facilitate accurate knowledge transfer. Furthermore, technological tools, including e-learning platforms and mobile applications, can enhance teachers' ability to incorporate IKS effectively. Providing incentives, such as scholarships for IKS specialization and funding for innovative classroom projects, along with mentorship programs, can further support teachers. Governments and educational institutions must back these initiatives by mandating the inclusion of IKS in teacher education curricula, allocating funding for resource development, and establishing policies that promote culturally inclusive education. By equipping teachers with the knowledge, skills, and resources needed, educational reform can become both culturally relevant and sustainable.

4.5 A Framework for Integration

Therefore, the need to fully incorporate IKS into advanced mathematical formula development will require a framework. This framework should include several key elements:

1. Cultural Relevance: Paying attention to the trove of cultural references to ensure that powerful mathematical ideas are set within the framing of indigenous Okoronkwo art, architecture and astronomy.
2. Collaborative Approach: Including indigenous knowledge custodians, teacher trainers, and mathematicians in designing and developing suitable indigenous material.
3. Pedagogical Flexibility: Teachers should be encouraged to take an open approach to integrating what is taught in the classroom with indigenous and Western ideas in mathematics that students can easily understand and appreciate.
4. Policy Support: Government curriculum development policies should provide finance for integrating IKS into the curriculum, training teachers to teach IKS insistently, and developing teaching aids that are pro-IKS.

5 CONCLUSION

Adopting Indigenous Knowledge Systems (IKS) into complex formula systems of technological and mathematical advancement is a potential area of curriculum development revolution in Southern Africa. The cultural background of the region and its traditions are a worthy starting point for developing means for improving the results of mathematics teaching and encouraging the emergence of new ideas in addressing modern problems. Implementation of IKS into the contents of Mathematics can help teachers close the gap between the already existing gaps between cultural practices and the Western way of teaching and learning Mathematics, thereby enhancing appreciation and understanding among the young generation.

It also helps to create a more intelligent school for the education of local communities and respect for modern society's practices while preparing the students for the globalization process. This fosters critical evaluation skills, creativity, and problem-solving skills backed by indigenous and Western chains of reasoning. Besides, such a curriculum can enable students to make positive use of the programme and apply mathematical knowledge in cultural and relevant contexts to effectively support their societies and communities in general.

However, to make this work, there is a need for the collective effort of educators, policymakers, community, and research so that IKS can be brought into its proper perspective. The former includes teacher training, curriculum reform, and research focusing on Indigenous mathematical practices, which are crucial in forming a framework. In the end, if Southern Africa embraces indigenous and modern mathematical knowledge, then Southern Africa can lead the way towards a fresh and entirely new Math Education which is culturally sensitive, to the benefit of producing a new generation of learners capable of solving world and regional problems.

6 METHODOLOGY

This study employed a qualitative and mixed-methods approach to explore the integration of Indigenous Knowledge Systems (IKS) into the development of advanced mathematical formulas, aiming to establish a framework for curriculum innovation in Southern Africa. The methodology was implemented as follows:

6.1 Research Design

A combination of exploratory and participatory action research (PAR) was utilized to engage local communities, educators, and curriculum developers in co-creating knowledge. This design ensured that indigenous perspectives were valued and systematically incorporated into advanced mathematical frameworks.

6.2 Data Collection Methods

Interviews and Focus Groups:
Semi-structured interviews and focus groups were conducted with indigenous knowledge holders (e.g., elders, cultural practitioners) and mathematicians to understand the principles and logic embedded in traditional systems.
Document Analysis:
Traditional texts, oral histories, and cultural artifacts relevant to indigenous mathematical practices were analyzed to uncover foundational principles and their applicability to modern mathematics.
Classroom Observations and Pilot Studies:
Observations were carried out in educational settings where experimental curricula incorporating IKS were implemented. This approach assessed the practical challenges and effectiveness of such frameworks.

6.3 Framework Development

Using grounded theory, the collected data were coded and analyzed to develop a theoretical framework. This framework outlined processes for:
Identifying indigenous concepts compatible with mathematical principles.
Adapting these concepts into advanced mathematical problem-solving and formula development.
Embedding this integration into Southern Africa's curriculum.

6.4 Validation and Testing

Workshops and Expert Panels:
Iterative workshops with curriculum experts, educators, and indigenous representatives were conducted to validate the developed framework.
Case Studies and Pilot Programs:
Specific schools or districts in Southern Africa were selected as case studies to implement and test the proposed framework in real educational contexts.

6.5 Data Analysis

Data from interviews, focus groups, and pilot programs were analyzed thematically to identify recurring patterns and insights. Quantitative data from pilot studies, such as student performance metrics, were statistically analyzed to measure outcomes.

6.6 Ethical Considerations

Ethical protocols were adhered to throughout the study, including obtaining informed consent from participants, respecting indigenous intellectual property rights, and collaborating with community leaders to ensure cultural sensitivity.

6.7 Outcome

The study successfully produced a validated framework for integrating IKS into mathematical formula development, contributing to curriculum innovation and promoting culturally relevant education in Southern Africa.

7 THEORETICAL FRAMEWORK

The theoretical framework for this study was grounded in Constructivist Learning Theory and Ethnomathematics Theory, which provided a foundation for integrating Indigenous Knowledge Systems (IKS) into advanced mathematical formula development and curriculum innovation in Southern Africa. Constructivist Learning Theory emphasizes that learners construct knowledge through active engagement and interaction with their environment. This study utilized the theory to highlight how indigenous knowledge, rooted in cultural practices and real-life experiences, could serve as a foundation for understanding and applying advanced mathematical concepts. By situating mathematical learning within culturally relevant contexts, the framework encouraged learners to develop meaningful connections between abstract formulas and practical applications.

Additionally, Ethnomathematics Theory, which explores the relationship between mathematics and culture, offered critical insights into identifying and validating indigenous mathematical practices, such as patterns, measurements, and problem-solving techniques embedded in traditional African contexts. This theory supported the recognition of IKS as a legitimate and valuable source of mathematical knowledge, enriching modern curricula and promoting inclusivity. Together, these frameworks informed the study’s design, data collection, and analysis, ensuring that the proposed curriculum innovation was both culturally responsive and pedagogically sound.

8 DISCUSSION

Integrating indigenous culture with modern educational content and teaching methods is a crucial step toward making educational reform more culturally adaptable and sustainable. This study highlights that the effective fusion of Indigenous Knowledge Systems (IKS) with contemporary education requires a nuanced understanding of both the cultural context and the demands of modern educational frameworks. Key considerations for achieving this integration are discussed below

Firstly, curriculum design must embrace contextual relevance by incorporating indigenous cultural practices, values, and problem-solving techniques into modern subjects like mathematics, science, and social studies. For example, traditional methods of measurement, navigation, or agricultural practices can serve as a foundation for teaching advanced concepts. This approach not only enriches the learning experience but also instills a sense of pride and identity among learners by validating their cultural heritage.

Secondly, teacher training and capacity building are essential for this integration. Educators must be equipped with the knowledge and skills to blend indigenous knowledge with modern teaching methods effectively. Professional development programs should include workshops on IKS and culturally responsive pedagogy to empower teachers to bridge the gap between traditional and contemporary knowledge systems.

Thirdly, community involvement is pivotal. Indigenous elders, cultural practitioners, and local leaders should be actively engaged in the educational process to ensure the authenticity and accuracy of cultural content. Their participation fosters a collaborative approach, where education becomes a shared responsibility, enhancing its relevance and acceptance within the community.

Furthermore, innovative teaching methods such as project-based learning, storytelling, and experiential learning can help integrate indigenous culture into classroom practices. These methods allow students to apply their cultural knowledge in solving real-world problems, making learning more practical and meaningful.

Lastly, the sustainability of such reforms depends on policy support and resource allocation. Governments and educational authorities must institutionalize the integration of IKS through supportive policies, adequate funding, and the development of culturally relevant teaching materials.

In conclusion, the integration of indigenous culture with modern education requires a holistic approach that respects and values cultural diversity while addressing global educational standards. By ensuring contextual relevance, empowering educators, involving communities, and adopting innovative methods, educational reform can become culturally adaptable and sustainable, ultimately fostering inclusive and transformative learning environments.

Policy support is essential for promoting the integration of Indigenous Knowledge Systems (IKS) into modern education, as it provides the framework, resources, and accountability needed to ensure this integration is systematic, consistent, and sustainable. Policies lend legitimacy to IKS by embedding it into national education standards, mandating its inclusion in curricula, and emphasizing its value in fostering culturally relevant education. Effective policies ensure funding for critical areas such as teacher training, resource development, and community engagement, addressing key barriers to implementation. For successful policy formulation, steps such as conducting stakeholder consultations, mapping indigenous practices, and drafting inclusive policies are critical. These steps should involve educators, indigenous knowledge holders, policymakers, and other stakeholders to ensure relevance and inclusivity. During implementation, capacity building for teachers, the development of culturally relevant resources, and the establishment of support systems are crucial to facilitate the integration process. Monitoring and evaluation play a vital role in assessing policy effectiveness, using tools like surveys, interviews, and performance metrics to gather feedback and refine strategies. By ensuring comprehensive policy formulation, implementation, and evaluation, educational reforms can sustainably integrate IKS, making education systems more culturally adaptable, inclusive, and globally relevant.

9 RECOMMENDATION

The possibility of bridging IKS and advanced mathematical formula formation is an innovative educational opportunity nicely situated in Southern Africa. The following recommendations can guide curriculum experts in shaping the future of education in this regard:

Acknowledge Indigenous Knowledge as a Valid System of Thought: First and foremost, there is a dire need to appreciate the IK systems for what they are, that is, as systems that exist in parallel with what has conventionally been characterized as scientific and mathematical systems. It is important to recognize the complexity of mathematics embedded in Indigenous societies, such as the use of geometrical aspects in Indigenous arts and buildings and even the allocation of resources.

Curriculum Framework Development: Establish a sound guided framework that integrates IKS into the curriculum at different levels of education. This may include using Indigenous elders, university scholars, and mathematicians to employ and develop cross-cultural education to intersect the Math concept with pre-existing knowledge. While developing the curriculum, it is possible to make sure that besides studying IKS as one of the important components of the African people and their culture, the students will also learn the newest methods to solve different logical tasks in mathematics.

Interdisciplinary Approach: Promote solutions where mathematics and quantitative processes integrate with culture, history, and anthropology. This would allow students to understand how indigenous systems have developed their calculation methods in Agriculture, Astronomy, and ecology.

Pedagogical Methods: Use other culturally sensitive approaches to teaching/learning that are sensitive to Indigenous people's knowledge. This could involve folktales, narrative histories, and uses of numbers drawn from the contexts of real life. Involve the students in activity and discovery-based learning practices that will seek to relate IKS to complicated mathematical concepts.

Research and Development: Call for more studies concerning mathematics content in Indigenous communities. It is possible to explore more concepts in Indigenous people for various purposes, e.g., geolocation, measurement of time, or management of resources, or to analyze concrete solutions discovered by natives. This may help formulate new mathematical theories or even higher mathematical formulae that embody Indigenous intelligence.

Collaborations and Partnerships: The following relationships should be developed: uni-collaboration-university: The university should seek partnerships with local Indigenous peoples and relevant government departments to finance the promotion of IKS in tertiary education. These could also enhance the creation of capacity-building activities for teachers to help them implement IKS practices in their learning institutions.

Evaluation and Feedback: These are the ways in which any mechanism necessary for evaluation and feedback during the integration of IKS into the curriculum is put in place. This includes feedback from the Indigenous people, students, and educators to make the integration progress more reverent, authentic, and useful to everybody.

Global Engagement and Adaptability: Last, localize the integration of IKS into mathematical curricula as a worldwide endeavour. Promote ST philanthropy by ensuring Southern African institutions expose their experiences with other institutions globally, thus widening the depth of mathematics.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Odora-Hoppers C A. Indigenous Knowledge and the Integration of Knowledge Systems in South Africa, 2002.
[2] Battiste M. Indigenous Knowledge and Pedagogy in First Nations Education: A Literature Review with Recommendations, 2002.
[3] Ascher M.Mathematics and Culture, 1991.
[4] Vithal R, Skovsmose O. The Role of Mathematics in Social and Cultural Practices, 1997.
[5] Gerdes P. Ethnomathematics: A Multicultural View of Mathematical Ideas, 1999.
[6] Kline M. Mathematics in Western Culture,1972.
[7] Knorr W. The Ancient Tradition of Geometrical Problems, 1993.

[8] Lindberg D C. The Beginnings of Western Science (2nd ed.), 2007.
[9] Weinstein A. Calculus: A History of the Mathematics of Change, 2012.
[10] Smith L T. Decolonizing Methodologies: Research and Indigenous Peoples, 1999.
[11] Kriel D. Indigenous knowledge systems and their influence on African mathematical practices. Mathematics Education Journal, 2012, 48(2): 112-125.
[12] Maphosa D. Reclaiming indigenous knowledge systems in the context of Southern African mathematics education. African Journal of Indigenous Knowledge, 2019, 7(1): 33-47.
[13] Bashir M, Smith D. Decolonizing education: Integrating indigenous knowledge systems into mathematics curricula. Journal of Educational Reform, 2017, 12(4): 234-249.
[14] Nhemachena A, Jansen J. Colonial education and its impact on mathematics education in Southern Africa. International Journal of Educational Development, 2020, 52: 19-32.
[15] Chikoko V, Jere M, Mberwa B. Embracing modern mathematics within African mathematics knowledge: A need for curriculum transformation, 2015.

Multidisciplinary Journal of Engineering and Technology
ISSN: 3007-6900
DOI: https://doi.org/10.61784/mjet3020

RESOURCE SCHEDULING STRATEGY FOR INTERNET OF THINGS EDGE COMPUTING BASED ON ARTIFICIAL POTENTIAL FIELD ALGORITHM

ZongYang Du
Kwangwoon University, Seoul, South Korea.
Corresponding Email: duzy@kw.ac.kr

Abstract: With the rapid development of the mobile Internet, smart devices are playing an increasingly important role in people's lives. Edge computing is a key technology for solving resource scheduling problems and is of great significance for meeting high-performance user requirements. This paper deeply studies the resource scheduling strategy for Internet of Things edge computing based on artificial potential field algorithm. It analyzes the architecture of edge computing, resource scheduling methods, research problems, and technical methods, and discusses its application scenarios and challenges. Through the introduction of artificial potential field algorithm, the edge computing resource scheduling is optimized to improve system performance and user experience.
Keywords: User experience; System performance; Internet of things; Key technology

1 INTRODUCTION

With the rapid development of the Internet of Things, the data generated by smart devices is massive, which puts forward higher requirements for resource capacity and processing capabilities. The centralized processing mode of cloud computing has limitations in terms of bandwidth, latency, and energy consumption, making it difficult to meet users' high-performance requirements. Edge computing moves services and functions to the user's side, providing users with communication, storage, and computing capabilities. Resource scheduling is the core issue of edge computing and the key to optimizing resource allocation and improving system performance.

2 EDGE COMPUTING ARCHITECTURE

2.1 User Layer

The user layer mainly consists of IoT devices, including drones, networked autonomous vehicles, AR devices, public security surveillance cameras, medical sensors, IoT devices in intelligent manufacturing, smart anti-theft detectors, etc. These devices have sensing capabilities and also possess certain storage and computing capabilities [1]. They can collect various data, process it, and upload it as input for application services (Figure 1).

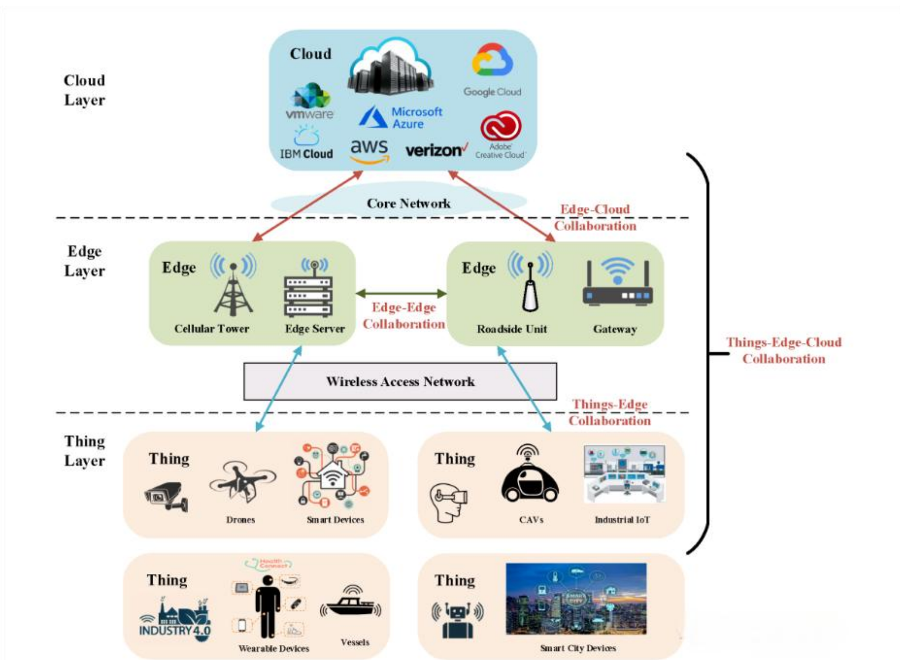


Figure 1 Edge Computing Resource Scheduling Framework

© By the Author(s) 2025, under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0).

2.2 Edge Layer

The edge layer is widely distributed between the end devices and the cloud computing center and is deployed together with the access points, such as gateway routers, RSU(Road Side Unit) [2], base stations, etc. Edge nodes have strong storage and computing capabilities. They receive, process, and forward the data streams from the user layer, providing services such as intelligent perception, security and privacy protection, data analysis, intelligent computing, process optimization, and real-time control. The edge layer also includes edge manager software, which can perform operations on edge computing nodes through business orchestration or direct invocation to complete tasks.

2.3 Cloud Computing Layer

The cloud computing layer consists of powerful computing centers and storage units. Tasks that cannot be processed by the edge layer and tasks involving comprehensive global information still need to be completed in the cloud computing center. The cloud computing layer provides decision support systems and application service programs in specific fields such as intelligent production [3], networked collaboration, service extension, and personalized customization. The cloud computing layer receives data streams from the edge layer and sends control information to the edge layer and through the edge layer to the user layer, optimizing resource scheduling and on-site production processes from a global perspective (Figure 2).

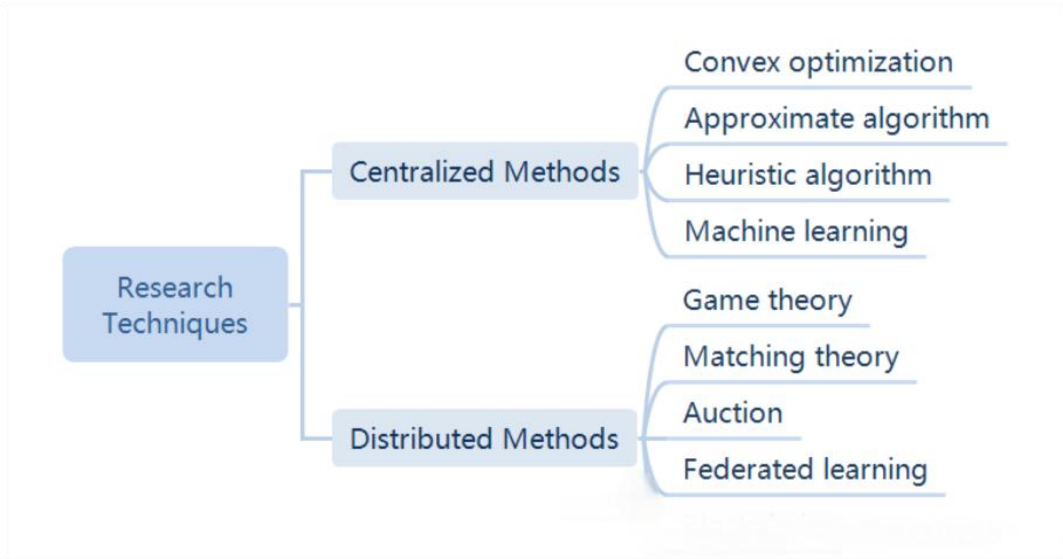


Figure 2 Edge Computing Resource Scheduling Method

3 RESOURCE SCHEDULING COLLABORATION MODES

3.1 "User-Edge" Collaboration

Under this collaboration mode, a close connection is established between users and the edge. Based on their own needs, users offload tasks to edge nodes. Relying on their powerful computing capabilities and abundant resources, edge nodes provide users with efficient processing services [4]. For example, in the smart home scenario, users' smart devices transmit various collected data to edge nodes, and the edge nodes analyze and process the data, such as intelligent control of the indoor environment and equipment status monitoring. In this process, in addition to completing data calculations, edge nodes can also provide real-time feedback to users to ensure that they can have a good experience.

3.2 "User-Edge-Cloud" Collaboration

When users offload tasks to edge nodes, the edge nodes will conduct preliminary processing on the tasks. When encountering complex tasks or when further resource support is needed, the edge nodes will send part of the tasks or data to the cloud center. Relying on its powerful computing capabilities and vast storage resources, the cloud center will conduct in-depth analysis and processing on these tasks. For example, in the field of intelligent healthcare, edge nodes collect patients' health data, which is then transferred to the cloud center after preliminary processing. The cloud center conducts a more comprehensive analysis and diagnosis and then returns the results to the edge nodes [5]. Finally, the edge nodes feed back the results to the users. This collaboration mode realizes the complementary advantages of the edge and the cloud center, providing users with more comprehensive and high-quality services.

3.3 "Edge-Edge" Collaboration

Edge nodes collaborate with each other, share resources, and jointly handle tasks. In this process, each edge node gives play to its respective advantages to achieve optimized resource allocation. For example, in the intelligent transportation system, different edge nodes collect and process traffic data for different road sections. When a node encounters congestion, other nodes can share resources and jointly handle the situation to help this node relieve the congestion. This collaboration mode improves the overall performance and resource utilization rate of edge nodes and ensures the stable operation of the entire system [6].

3.4 "Edge-Cloud" Collaboration

Edge nodes and the cloud center conduct resource scheduling. Edge nodes send tasks to the cloud center, and the cloud center will process the tasks and return the results. In this process, the cloud center provides powerful computing support and resource guarantee for edge nodes [7]. For example, in industrial production, edge nodes send production data to the cloud center for analysis and processing. Based on the analysis results, the cloud center provides optimization suggestions and decision-making support for edge nodes. This collaboration mode enables edge nodes to play their roles better and improves the efficiency and quality of the entire system (Figure 3).

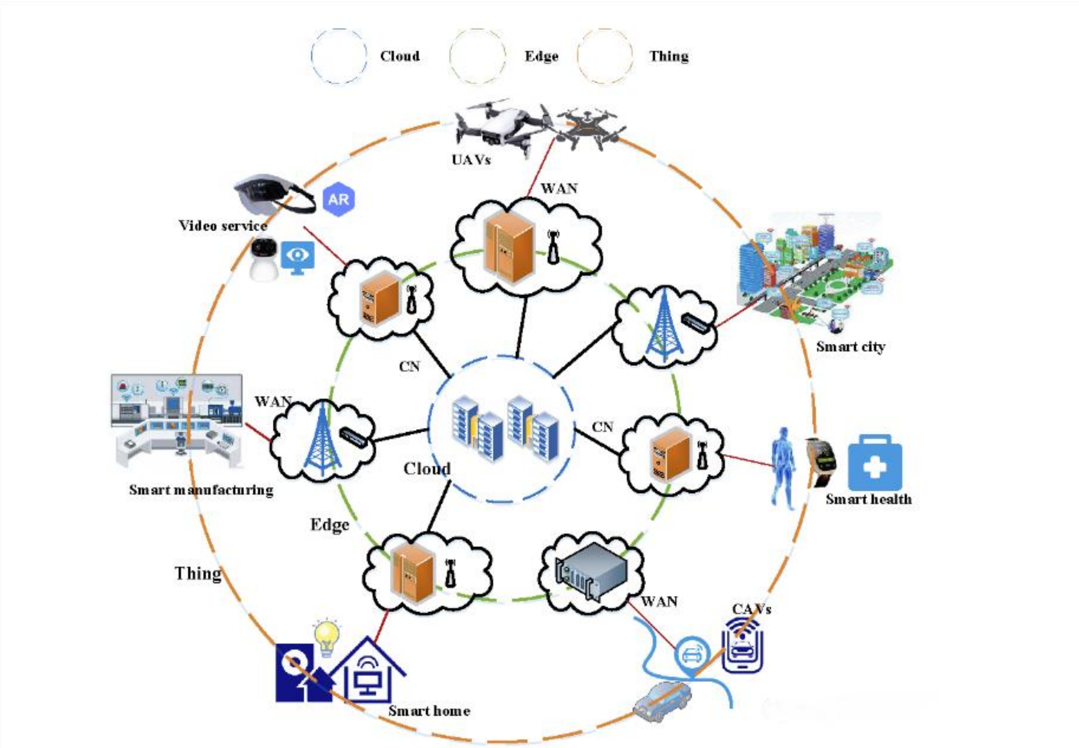


Figure 3 Main Application Scenarios in Edge Computing Resource Scheduling

4 RESEARCH PROBLEMS IN RESOURCE SCHEDULING

4.1 Computational Offloading

Computational offloading is an important part of edge computing resource scheduling, which refers to achieving specific goals through the transfer of computational tasks among different layers. In practical applications, there are multiple levels of interaction in this process. In terms of offloading directions, user-to-edge offloading is the most common way. Users offload their computational tasks to edge nodes and use the computing resources and processing capabilities of edge nodes to complete them [8]. For example, various applications run on mobile devices. When users need to process a large amount of data or perform complex calculations, offloading tasks to edge nodes can greatly reduce latency. Edge-to-edge offloading refers to the transfer of tasks between edge nodes. When an edge node has a heavy load or encounters specific tasks, it can transfer some tasks to other edge nodes to balance the load and improve resource utilization. Cloud-to-edge offloading means that the cloud center offloads tasks to edge nodes. The cloud center has powerful computing capabilities and resources [9]. By offloading tasks to edge nodes, the cloud center can make better use of edge resources and achieve more efficient computing and services. In terms of offloading granularity, full offloading means completely transferring the entire task from one layer to another. This method is suitable for some scenarios that require high computing resources and where tasks are relatively independent. For example, for large data processing tasks, the entire task can be offloaded to edge nodes for processing. While partial offloading is to transfer a part of the task, which is more common in some complex tasks. For example, in a video processing task, the processing tasks of some video frames can be offloaded to edge nodes, and the other parts are processed by the cloud center.

4.2 Resource Allocation

Resource allocation is a key link to ensure the effective utilization of resources in the edge computing system. In the edge computing environment, the rational allocation of resources directly affects the processing efficiency of tasks and system performance. Computing resources are a key part of it. It includes devices such as processors and memory. Computing resources need to be allocated according to the requirements of tasks and the capabilities of edge nodes. For example, for some tasks that consume a large amount of computing resources, such as artificial intelligence model training, sufficient computing resources need to be allocated to edge nodes to ensure the smooth execution of tasks. Communication resources are also an important part of resource allocation [10]. It includes network bandwidth, signal transmission, etc. In the edge computing system, the allocation of communication resources determines the transmission speed and quality of data. For example, in the communication between Internet of Things devices, sufficient network bandwidth is needed to ensure the timely transmission of data. Storage resources are for the storage and management of data. Edge nodes need to store a large amount of task data and intermediate results. The rational allocation of storage resources can ensure the security and effective management of data.

4.3 Resource Configuration

At the service provider level, resource configuration is an active and strategic process. The decentralization of cloud services to the edge means that some services and functions of the cloud center are migrated to edge nodes for the optimization of resource configuration. These are beneficial to the improvement of the service capabilities and response speeds of edge nodes. The optimal deployment of edge servers involves reasonably arranging the locations and performances of servers according to the actual needs and environmental characteristics of edge nodes. For example, the optimized deployment of edge servers in some remote areas or places with poor network conditions can improve the service quality of edge nodes [11]. The allocation of the number of edge resources is also an important link. Resources should be reasonably allocated according to the requirements of tasks and the carrying capacities of edge nodes. Virtual edge resource configuration refers to creating virtual resources according to the needs of different tasks. From the user's perspective, resource configuration is a passive process. It is about the optimal allocation and matching scheme between user tasks and edge resources. For example, when users use edge computing services, they hope to obtain the best resource configuration to meet their own needs. This requires reasonable matching according to the characteristics of user tasks and the status of edge resources. When the task requirements of users change, the resource configuration should also change accordingly [12]. This passive resource configuration method ensures that users can obtain the most suitable resources and services in the edge computing environment.

5 RESOURCE SCHEDULING STRATEGIES BASED ON ARTIFICIAL POTENTIAL FIELD ALGORITHM

5. 1 Principle of Artificial Potential Field Algorithm

The artificial potential field algorithm is simulated from the interaction of objects in a physical field. Its core idea is to describe the state and behavior of objects based on the construction of potential field functions. The construction of potential field functions is particularly important in the context of edge computing resource scheduling. It takes the state of resources and the needs of users as important bases and incorporates factors such as the load situation of resources and the distance from users. For example, for the resources of edge nodes, when the resource load is relatively large, in order to avoid excessive concentration of resources, a potential field function can be established to generate a repulsive force among resources, thus dispersing the load [13]. Meanwhile, if an edge node is relatively close to the user, considering the user's need for a quick response, we can construct a potential field function to generate an attractive force between the resources and the user. This potential field function, which is constructed based on the resource state and user needs, can provide an effective mathematical model for resource scheduling [14].

5.2 Algorithm Implementation Steps

In the process of edge computing resource scheduling, first, determine the parameters of the potential field function according to the initial state of edge computing resources [15]. The parameters of the potential field function directly affect its characteristics. For example, the initial load situation of resources, the distance between resources and users, etc. Then calculate the potential field force of each resource according to the potential field function. This potential field force can be manifested as the attractive or repulsive force between resources.

For example, when there is a load difference between resources, the potential field force will make resources move from high-load areas to low-load areas, so that resources can be evenly allocated. Then, adjust the state of resources according to the magnitude and direction of the potential field force [16]. This includes aspects such as resource allocation and scheduling. For example, when resources have an attractive force, they are scheduled to users, and when resources have a repulsive force, they are also scheduled to users. Finally, repeat the above steps until the purpose of resource scheduling is achieved. This is a dynamic and continuous process, and the potential field function as well as the resource state will be continuously adjusted with the changes in the resource state and user needs.

5.3 Advantages and Application Scenarios

The artificial potential field algorithm has great advantages in edge computing resource scheduling. Firstly, it can quickly respond to user needs [17]. The potential field function can be quickly adjusted when user needs change to achieve efficient resource allocation. For example, during the driving process of intelligent networked vehicles, when the user's needs change, the potential field algorithm can timely adjust resource allocation to ensure the safe driving of vehicles. Secondly, this algorithm can be adjusted in real time according to the dynamic changes of resources. The state of edge computing resources is constantly changing, and the potential field algorithm can monitor these changes in real time and adjust resource allocation accordingly. For example, in video services, as the video content changes and the number of user needs increases, the potential field algorithm can adjust resource configuration in real time to ensure the smooth playback of videos. In addition, the artificial potential field algorithm has good adaptability to complex environments. Whether in the complex urban environment or in the complex scenarios of industrial manufacturing, the algorithm can effectively play its role. For example, in the construction of a smart city, the potential field algorithm can optimize resource scheduling according to the city's resource distribution and user needs [18-19]. It has a wide range of application scenarios, involving drones, intelligent networked vehicles, video services, smart cities, smart health, intelligent industrial manufacturing, smart homes, etc. The potential field algorithm can optimize resource scheduling according to the flight state and mission requirements of drones during their flight process [20]. The potential field algorithm can also reasonably allocate resources according to the user needs and resource status in the smart home scenario.

6 CONCLUSION

This paper has deeply studied the resource scheduling strategy for Internet of Things edge computing based on artificial potential field algorithm. Through the edge computing architecture, resource scheduling collaboration methods, research problems, and technical methods, the application of artificial potential field algorithm in resource scheduling has been analyzed. The artificial potential field algorithm can effectively optimize resource scheduling and improve system performance and user experience. However, in practical applications, there are still many problems that need to be studied and solved. The future development of edge computing technology will provide more opportunities and challenges for resource scheduling in Internet of Things edge computing. We hope that more scholars will engage in the research of edge computing resource scheduling to promote the development and application of related technologies.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCE

[1] Hari Prasad Acharja, Sonam Choki, Dorji Wangmo, et al. Development of fog visibility enhancement and alert system using IoT. Cogent Engineering, 2024, 11(1).
[2] Amenu Leta Duguma, Xiuguang Bai. How the internet of things technology improves agricultural efficiency. Artificial Intelligence Review, 2024, 58(2): 63-63.
[3] Basil C E Oguguo, Barnabas C Madu, Anayo David Nnaji, et al. The Predictive Potency of Internet of Things Cybersecurity Technology on Classroom Assessment Practices in Higher Institutions of Learning in Nigeria. SN Computer Science, 2024, 5(8): 1174-1174.
[4] Lingya Zhang. Internet of things vs. factory of things:an evaluation of evolving technologies for corporate sustainable development. International Journal of Electronic Business, 2025, 20(1): 17-33.
[5] Mst Nargis Aktar, Nilanjana Basak, Shuvo Biswas, et al. SPR-Based Fiber Optic Sensor for the Development of Internet of Things (IoT) Technologies. Plasmonics, 2024, (prepublish): 1-11.
[6] Apinan Aurasopon, Thawatchai Thongleam, Sanya Kuankid. Integration of IoT Technology in Hydroponic Systems for Enhanced Efficiency and Productivity in Small-Scale Farming. Acta Technologica Agriculturae, 2024, 27(4): 203-211.
[7] Neerav Sharma, Shubham Bhattacharjee, Rahul Dev Garg, et al. Sustainable management and agriculture resource technology system using remote sensing descriptors and IoT. Geomatica, 2024, 76(2): 100040-100040.
[8] Rehman Shafique Ur, Usman Muhammad, Fernando Yudi, et al. Improving manufacturing supply chain performance:nexus of industrial Internet of Things, blockchain technology and innovativeness. Journal of Science and Technology Policy Management, 2024, 15(6): 1641-1664.
[9] Boyuan Wang, Xiali Shi, Xihao Han, et al. The digital transformation of nursing practice:an analysis of advanced IoT technologies and smart nursing systems. Frontiers in Medicine, 2024, 11: 1471527-1471527.
[10] Ding Dawei, Shankar Achyut. Industrial mechanical equipment fault detection and high-performance data analysis technology based on the Internet of Things. Intelligent Decision Technologies, 2024, 18(4): 3171-3184.

[11] Zhang Zongju, Shankar Achyut. Design and simulation of electronic communication positioning system based on Internet of Things technology. Intelligent Decision Technologies, 2024, 18(4): 3347-3363.
[12] El Jaouhari Asmae, Arif Jabir, Samadhiya Ashutosh, et al. An environmental-based perspective framework:integrating IoT technology into a sustainable automotive supply chain. Benchmarking: An International Journal, 2024, 31(10): 3655-3689.
[13] Valentin Iordache, Marius Minea, Răzvan Andrei Gheorghiu, et al. Integrating Connected Vehicles into IoT Ecosystems:A Comparative Study of Low-Power, Long-Range Communication Technologies. Sensors, 2024, 24(23): 7607-7607.
[14] IoT MVNO Market Growing Trends, Future Outlook, Advance Technology, Global Size, Share And Forecast-2028. M2 Presswire, 2024,
[15] Liying Zhao, Lina Ding. Wireless sensor network based on IoT automation technology application in green supply chain management of automobile manufacturing industry. The International Journal of Advanced Manufacturing Technology, 2024, (prepublish):1-12.
[16] Seetah Almarri, Ahmed Aljughaiman. Blockchain Technology for IoT Security and Trust: A Comprehensive SLR. Sustainability, 2024, 16(23): 10177-10177.
[17] Xueting Yang, Bing Sun, Shilong Liu. Study of technology communities and dominant technology lock-in in the Internet of Things domain-Based on social network analysis of patent network. Information Processing and Management, 2025, 62(1): 103959-103959.
[18] Yasir Ali, Habib Ullah Khan, Faheem Khan, et al. Building integrated assessment model for IoT technology deployment in the Industry 4. 0. Journal of Cloud Computing, 2024, 13(1): 155-155.
[19] Agostino Marengo, Alessandro Pagano, Vito Santamato. An efficient cardiovascular disease prediction model through AI-driven IoT technology. Computers in Biology and Medicine, 2024, 183: 109330-109330.
[20] Ruoqian Yang, Xingfang Peng. Blockchain Technology Marketing Sustainable and Accurate Push System Based on Big Data and Internet of Things. Advances in Computer, Signals and Systems, 2024, 8(7).

Multidisciplinary Journal of Engineering and Technology
ISSN: 3007-6900
DOI: https://doi.org/10.61784/mjet3021

THERMAL EFFECTS ON THE LD END-FACE PUMPED ND:YAG LASER MEDIUM

XingYu Liu*, JianSheng Zhang, Wen Du
School of Basic Science, Xi 'an Technological University, Xi 'an 710021, Shaanxi, China.
Corresponding Author: XingYu Liu, Email: a2210211003@163.com

Abstract: Solid-state lasers are widely used in various industries due to their high stability, compact structure, and excellent beam quality. However, thermal effects have long been a significant constraint on improving laser performance. Thermal effects are primarily caused by the heat deposition resulting from the absorption of pump energy by the laser medium in solid-state lasers. Studying the heat conduction equation of the laser medium is crucial for understanding thermal effects. This paper provides a detailed theoretical derivation for calculating the heat conduction equation of the laser medium in solid-state lasers. A series expansion is used to derive the analytical expression for the heat conduction equation, and the fourth-order Runge-Kutta method is employed to numerically solve it. The temperature simulation results of the laser medium end face using both methods are compared to verify the correctness of the analytical derivation process and to analyze the causes of any errors.
Keywords: Heat conduction equation; Power series; Runge-Kutta; Temperature simulation

1 INTRODUCTION

In recent years, solid-state lasers based on Nd:YAG crystals have gained wide applications due to their excellent beam quality, high stability, and compact design. These lasers are commonly used in fields such as material processing, medical treatments, and scientific research. However, one of the key challenges in the performance of these lasers is the thermal effects caused by the pumping process. A series of effects caused by the change in the end-face temperature gradient are referred to as the thermal effects of the laser [1-3]. When the laser diode (LD) pumps the Nd:YAG medium, the energy absorbed from the pump light generates heat, leading to a temperature gradient within the laser medium. This temperature gradient can severely affect the laser's efficiency, beam quality, and overall performance, causing phenomena such as thermal lensing, which distorts the laser beam and reduces output power.
Understanding the thermal effects in Nd:YAG laser media is crucial for optimizing laser design and improving its efficiency. As the heat generated by the absorbed pump light is not evenly distributed, the resulting temperature gradient plays a critical role in determining the laser's performance. This study focuses on the thermal effects at the end face of the laser medium under LD pumping, simulating the temperature distribution within the Nd:YAG crystal. The study provides an in-depth analysis of the temperature distribution at the medium's end face under different pump powers and pump beam radii.

2 HEAT CONDUCTION EQUATION OF THE LASER MEDIUM

The laser working substance is an important component of solid-state lasers. The Nd:YAG crystal is chosen for its high mechanical strength, good thermal conductivity, and low laser threshold, making it the subject of this study.

2.1 Establishing the Heat Conduction Equation

The temperature distribution inside the crystal is related to the heat power density generated by the pump light, the thermal properties of the crystal, the geometry, and external conditions [4-8]. Since the selected crystal has isotropic thermal properties and the cylindrical rod has axial symmetry, the steady-state equation and heat source equation in cylindrical coordinates are as follows:

$$\frac{\partial^2 T(r,z)}{\partial r^2}+\frac{1}{r}\frac{\partial T(r,z)}{\partial r}+\frac{\partial^2 T(r,z)}{\partial z^2}=-\frac{q(r,z)}{k} \tag{1}$$

$$\begin{cases} q(r,z)=\alpha\eta I_0 e^{[-2(\frac{r^2}{\omega^2})]}e^{(-\alpha z)} \\ I_0=\dfrac{P}{2\pi\int_0^\infty e^{-2\frac{r^2}{\omega^2}}rdr} \end{cases} \tag{2}$$

Where r and z are the radial and axial coordinates of the crystal, respectively, with the origin at the center of the pump face of the crystal, *T(r)* is the temperature, *q(r,z)* is the thermal power density function, PPP represents the pump power, η is the conversion efficiency (value of 0.3), k is the thermal conductivity of the crystal, which is 14 W/(m·K). and α is

© By the Author(s) 2025, under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0).

the absorption coefficient of the crystal (value of 0.3 cm^{-1}).
The heat variation along the direction of the laser rod's length z ultimately becomes **exp(−αz)**. Given this property, assuming the temperature distribution during the differential process is *T(r,z)=T(r) · T(z)=T(r)exp(−αz)*, we substitute it into equation (1) and obtain:

$$\frac{\partial^2 T(r)e^{-\alpha z}}{\partial r^2}+\frac{1}{r}\frac{\partial T(r)e^{-\alpha z}}{\partial r}+\alpha^2 T(r)e^{-\alpha z}+\frac{\alpha\eta I_0 e^{[-2(\frac{r^2}{\omega^2})]}e^{(-\alpha z)}}{k}=0 \quad (3)$$

Multiply both sides by exp(αz):

$$\frac{d^2T(r)}{dr^2}+\frac{1}{r}\frac{dT(r)}{dr}+\alpha^2 T(r)=-\frac{\alpha\eta I_0 e^{[-2(\frac{r^2}{\omega^2})]}}{k} \quad (4)$$

Set the constant B so that:

$$B=-\frac{\alpha\eta I_0}{k} \quad (5)$$

Now expand the right-hand side of (4) using a Taylor series:

$$\begin{aligned}&-\frac{\alpha\eta I_0}{k}e^{[-2(\frac{r^{2m}}{\omega^{2m}})]}\\&=B\sum_{n=0}^{\infty}\frac{1}{n!}(-2\frac{r^2}{\omega^2})^n\\&=\sum_{n=0}^{\infty}B\frac{(-2)^n}{\omega^{2n}*n!}r^{2n}\end{aligned} \quad (6)$$

Assume that *T(r)* is defined and can be analytically expressed within (r,r_0):

$$\begin{aligned}T(r)&=\sum_{n=0}^{\infty}A_n r^{2n}\\T'(r)&=\sum_{n=0}^{\infty}(2n)A_n r^{2n-1}\\T''(r)&=\sum_{n=0}^{\infty}(2n)(2n-1)A_n r^{2n-2}\end{aligned} \quad (7)$$

Substituting (7) into (4), get:

$$\begin{aligned}&\frac{d^2T(r)}{dr^2}+\frac{1}{r}\frac{dT(r)}{dr}+\alpha^2T(r)\\&=\sum_{n=0}^{\infty}(2n)(2n-1)A_n r^{2n-2}+r^{-1}\sum_{n=0}^{\infty}(2n)A_n r^{2n-1}+\alpha^2\sum_{n=0}^{\infty}A_n r^{2n}\\&=\sum_{n=0}^{\infty}(2n)(2n-1)A_n r^{2n-2}+\sum_{n=0}^{\infty}(2n)A_n r^{2n-2}+\alpha^2\sum_{n=0}^{\infty}A_n r^{2n}\\&=\sum_{n=0}^{\infty}(2n)^2A_n r^{2n-2}+\alpha^2\sum_{n=0}^{\infty}A_n r^{2n}\\&=\sum_{n=0}^{\infty}\left[(2n)^2A_n r^{2n-2}+\alpha^2A_n r^{2n}\right]\end{aligned} \quad (8)$$

From (4), (6), and (8), obtain:

$$\sum_{n=0}^{\infty}\left[(2n)^2A_n r^{2n-2}+\alpha^2A_n r^{2n}\right]=\sum_{n=0}^{\infty}B\frac{(-2)^n}{\omega^{2n}*(n)!}r^{2n} \quad (9)$$

Expanding both sides of (9) to the sixth term:

$$\begin{cases} n=0 & (0^2 A_0 r^{-2} + \alpha^2 A_0 r^0) \\ n=1 & (2^2 A_1 r^0 + \alpha^2 A_1 r^2) \\ n=2 & (4^2 A_2 r^2 + \alpha^2 A_2 r^4) \\ n=3 & (6^2 A_3 r^4 + \alpha^2 A_2 r^6) \\ n=4 & (8^2 A_4 r^6 + \alpha^2 A_2 r^8) \\ n=5 & (10^2 A_5 r^8 + \alpha^2 A_2 r^{10}) \end{cases} \tag{10}$$

Expanding the right-hand side of (9) to the sixth term:

$$\begin{cases} n=0 & B\dfrac{(-2)^0}{\omega^0 * 0!} r^0 \\ n=1 & B\dfrac{(-2)^1}{\omega^2 * 1!} r^2 \\ n=2 & B\dfrac{(-2)^2}{\omega^4 * 2!} r^4 \\ n=3 & B\dfrac{(-2)^3}{\omega^6 * 3!} r^6 \\ n=4 & B\dfrac{(-2)^4}{\omega^8 * 4!} r^8 \\ n=5 & B\dfrac{(-2)^5}{\omega^{10} * 5!} r^{10} \end{cases} \tag{11}$$

By examining equation (9), we find that the power terms of the variables on both sides are integer multiples of 2 starting from zero. Since both sides are equal, we conclude that the coefficients of the same powers of r on both sides are equal:

$$\begin{cases} r^0 & \alpha^2 A_0 + 2^2 A_1 = B\dfrac{(-2)^0}{0!*\omega^0} & A_1 = \left[B\dfrac{(-2)^0}{0!*\omega^0} - \alpha^2 A_0\right] * \dfrac{1}{4} \\ r^2 & \alpha^2 A_1 + 4^2 A_2 = B\dfrac{(-2)^1}{1!*\omega^2} & A_2 = \left[B\dfrac{(-2)^1}{1!*\omega^2} - \alpha^2 A_1\right] * \dfrac{1}{16} \\ r^4 & \alpha^2 A_2 + 6^2 A_3 = B\dfrac{(-2)^2}{2!*\omega^4} & A_3 = \left[B\dfrac{(-2)^2}{2!*\omega^4} - \alpha^2 A_2\right] * \dfrac{1}{36} \\ r^6 & \alpha^2 A_3 + 8^2 A_4 = B\dfrac{(-2)^3}{3!*\omega^6} & A_4 = \left[B\dfrac{(-2)^3}{3!*\omega^6} - \alpha^2 A_3\right] * \dfrac{1}{64} \\ r^8 & \alpha^2 A_4 + 10^2 A_5 = B\dfrac{(-2)^4}{4!*\omega^8} & A_5 = \left[B\dfrac{(-2)^4}{4!*\omega^8} - \alpha^2 A_4\right] * \dfrac{1}{100} \end{cases} \tag{12}$$

From equation (12), we obtain A_1 , A_2 , A_3 , A_4 , A_5 , and so on... Let A_0 be the first term of the equation, A_1 be the second term, A_2 be the third term, and so on. By this analogy, we can derive the inverse expression for A_n .

$$A_n = \frac{1}{(2n)^2}\left[B\frac{(-2)^{n-1}}{(n-1)!*\omega^{2n-2}} - \alpha^2 A_{n-1}\right] \tag{13}$$

Simplifying:

$$\begin{cases} T(r) = A_0 + \sum_{n=1}^{\infty} A_n r^{2n} \\ T(r) = A_0 + \sum_{n=1}^{\infty}\left[\dfrac{1}{(2n)^2}\left(B\dfrac{(-2)^{n-1}}{(n-1)!*\omega^{2n-2}} - \alpha^2 A_{(n-1)}\right)\right] r^{2n} \end{cases} \tag{14}$$

Substituting *T(r)* back into *T(r,z)*:

$$\begin{cases} T(r,z) = T(r)\exp(-\alpha z) \\ T(r,\ z) = A_0 \exp(-\alpha z) + \exp(-\alpha z)\sum_{n=1}^{\infty}\left[\frac{1}{(2n)^2}\left(B\frac{(-2)^{n-1}}{(n-1)!*\omega^{2n-2}} - \alpha^2 A_{(n-1)}\right)\right] r^{2n} \end{cases} \tag{15}$$

2.2 Boundary Conditions

In this study, the axial variation of the temperature gradient is neglected, and only the radial temperature distribution is considered. Since the heat transfer by convection from the end face to the air is minimal, the effect of convection on the temperature distribution at the end face is ignored. Since the crystal mainly dissipates heat through the side cooling system or cooling water, the side is assumed to be at a constant temperature. The boundary condition is thus determined as:

$$\begin{cases} \left.\frac{dT(r)}{dr}\right|_{r=0} = 0 \\ \left.T(r)\right|_{r=r_0} = T_0 \end{cases} \tag{16}$$

3 LASER MEDIUM END-FACE TEMPERATURE ANALYTICAL SIMULATION

Based on the analytical expression of the laser medium's heat conduction equation derived above, the temperature distribution at the laser medium's end face is simulated, while also exploring the impact of pump power and pump beam radius on the end-face temperature distribution.

3.1 Influence of Pump Power on Medium Temperature Distribution

The variation in pump power significantly affects the heat distribution at the laser medium's end face, leading to thermal lensing effects. Therefore, when using solid-state laser devices to achieve higher output power, it is important to minimize the thermal effects to improve the performance of the laser system. This requires studying the temperature distribution at the end face of the laser medium under different pump powers. Figure 1 shows the three-dimensional temperature distribution of the end face when the radius and length of the Nd:YAG rod are 0.2 cm and 0.5 cm, respectively, with a pump beam radius of 0.5 mm and pump powers of 5W, 10W, 15W, and 20W.

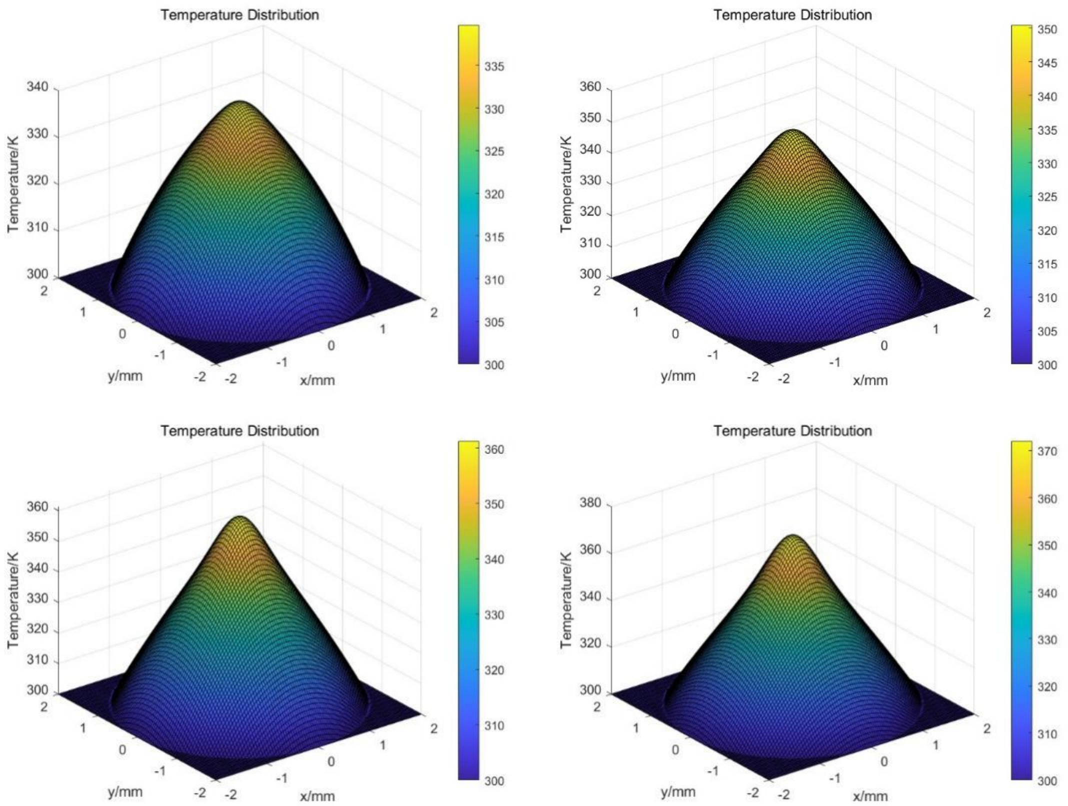


Figure 1 Temperature Distribution at the End-Face of the Laser Medium under Different Pump Powers

From Figure 1, it can be seen that the temperature at the center is the highest, and it increases as the pump power increases. The higher the pump power, the greater the amount of energy absorbed by the laser, leading to more heat accumulation at the smaller end face. The temperature distribution along the radial direction is shown in Figure 2.

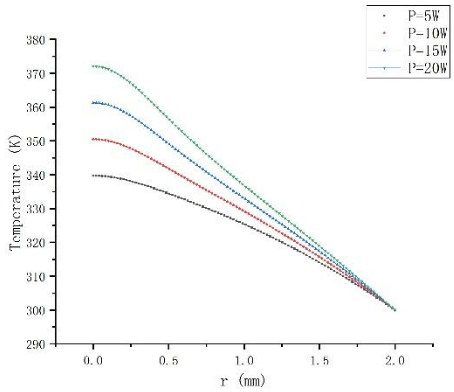


Figure 2 Radial Temperature Distribution of the Laser Medium under Different Pump Powers

Figure 2 shows the heat distribution at the end face of the laser medium under different pump powers. The temperature at the center of the laser medium is 339.73K, 350.52K, 361.32K, and 372.10K for pump powers of 5W, 10W, 15W, and 20W, respectively. As the pump power increases, the temperature at the center of the end face shows an increasing trend. Therefore, in order to ensure the output efficiency and meet the requirements during operation, the pump power should be appropriately reduced.

3.2 Influence of Pump Beam Radius on Medium Temperature Distribution

The change in pump beam radius has a certain impact on the heat source and the quality of the generated laser at the end face. Figure 3 shows the three-dimensional temperature distribution at the end face of the Nd:YAG rod with radii and lengths of 0.2 cm and 0.5 cm, respectively. The pump power is fixed at 10W, and the pump beam radii are 0.05 cm, 0.06 cm, 0.07 cm, and 0.08 cm.

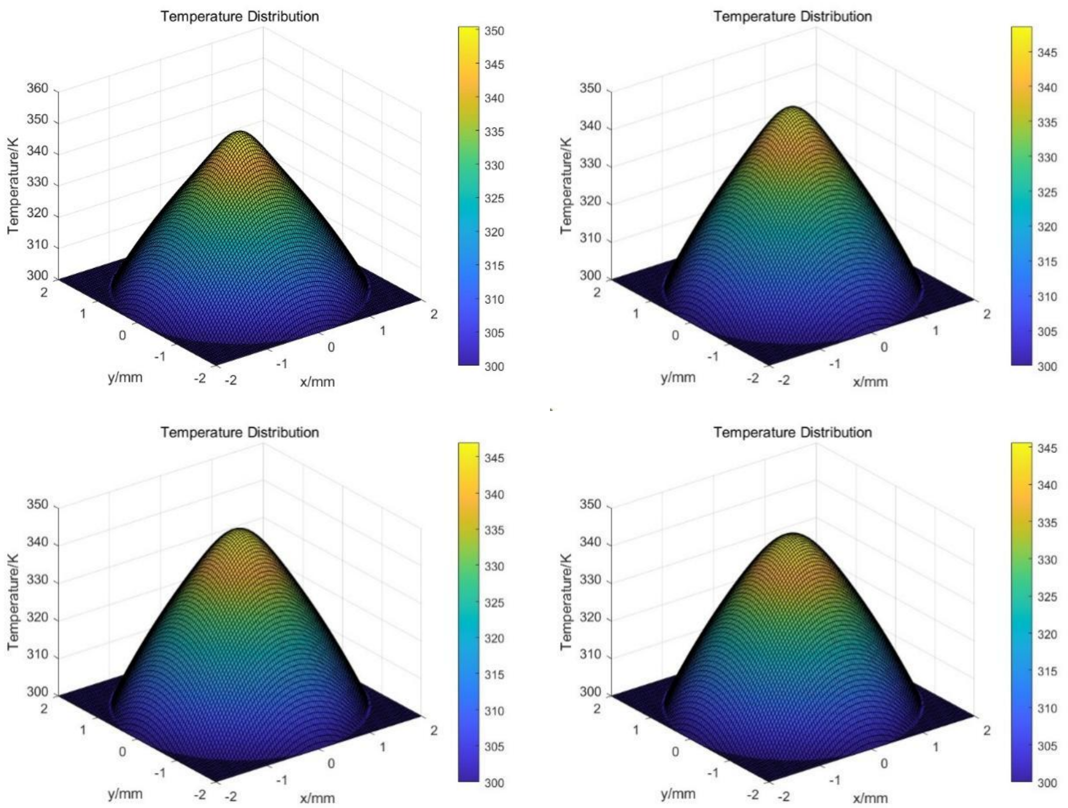


Figure 3 Temperature Distribution at the End-Face of the Laser Medium under Different Pump Beam Radii

From Figure 3, it can be seen that as the pump beam radius increases, the temperature at the center of the laser medium starts to decrease. This is mainly because the relationship between the pump beam radius and the heat power density is inversely proportional, so the heat accumulated at the laser medium's end face decreases. The specific temperature distribution along the radial direction is shown in Figure 4, which illustrates the heat distribution at the laser medium's end face under different pump beam radii.

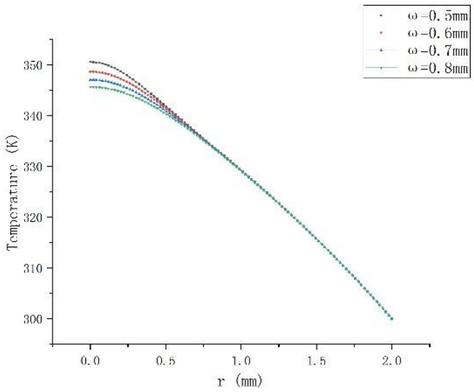


Figure 4 Radial Temperature Distribution of the Laser Medium under Different Pump Beam Radii

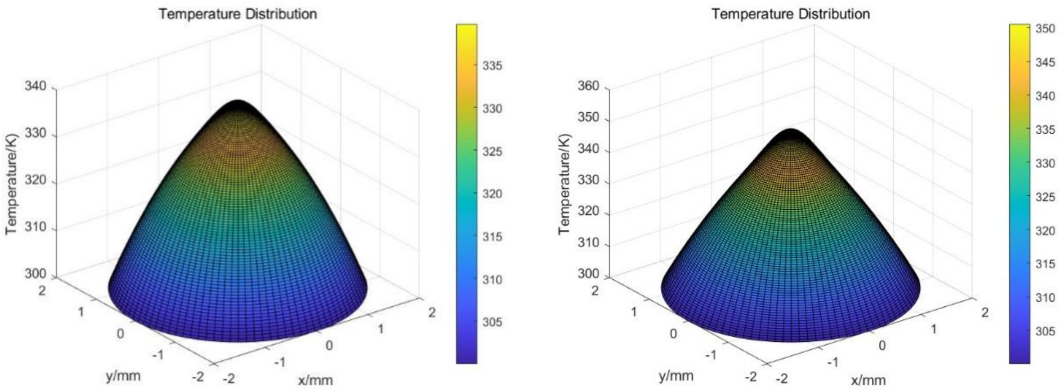
When the pump beam radius is 0.05 cm, 0.06 cm, 0.07 cm, and 0.08 cm, the center temperature of the laser medium end face is 350.52K, 348.63K, 347.03K, and 345.63K, respectively. After the pump beam radius decreases, the temperature at the end face decreases overall, but the decrease is very slow, with a temperature reduction controlled within 2K. Therefore, in practical applications, the choice of pump beam radius mainly depends on the requirements of the laser beam.

4 NUMERICAL SIMULATION OF THE LASER MEDIUM END-FACE TEMPERATURE

This section uses the fourth-order Runge-Kutta [9] method to numerically solve the heat conduction equation, simulate the temperature distribution at the laser medium's end face, and investigate the impact of pump power and pump beam radius on the end-face temperature distribution.

4.1 Influence of Pump Power on Medium Temperature Distribution

Figure 5 shows the three-dimensional temperature distribution at the laser medium's end face when the Nd:YAG rod has a radius and length of 0.2 cm and 0.5 cm, respectively, and the pump power is 5W, 10W, 15W, and 20W, with a pump beam radius of 0.5 mm.



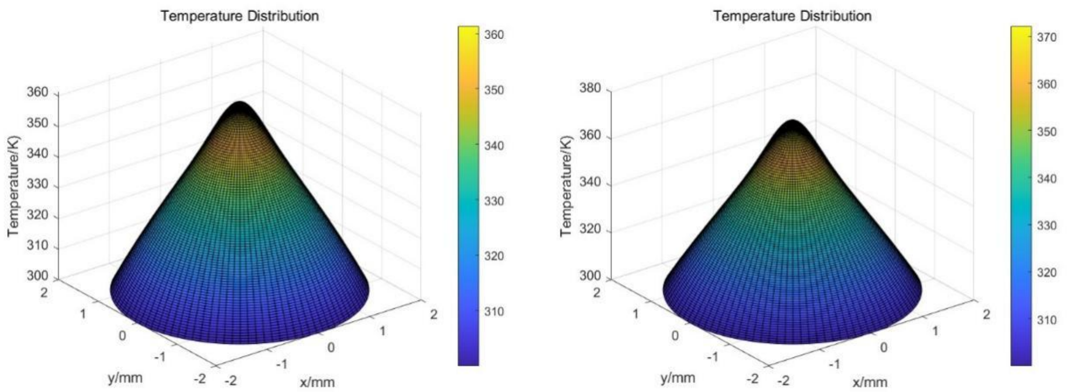


Figure 5: Temperature Distribution at the End-Face of the Laser Medium under Different Pump Powers

From Figure 5, it can be seen that along the radial direction, the temperature at the center is the highest, and it increases as the pump power increases. As the pump power increases, the energy absorbed by the laser medium also increases, leading to more heat accumulation at the smaller end face. The temperature distribution along the radial direction is shown in the figure 6.

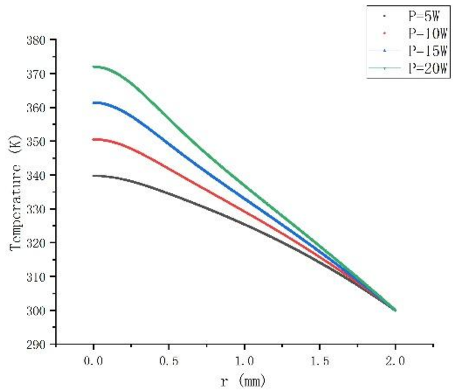


Figure 6 Radial Temperature Distribution of the Laser Medium under Different Pump Powers

Figure 6 shows the heat distribution at the laser medium's end face under different pump powers. The center temperature of the laser medium's end face is 339.74K, 350.53K, 361.32K, and 372.11K for pump powers of 5W, 10W, 15W, and 20W, respectively.

4.2 Influence of Pump Beam Radius on Medium Temperature Distribution

Figure 7 shows the three-dimensional temperature distribution at the laser medium's end face when the Nd:YAG rod has a radius and length of 0.2 cm and 0.5 cm, respectively. The pump power is fixed at 10W, and the pump beam radii are 0.05 cm, 0.06 cm, 0.07 cm, and 0.08 cm.

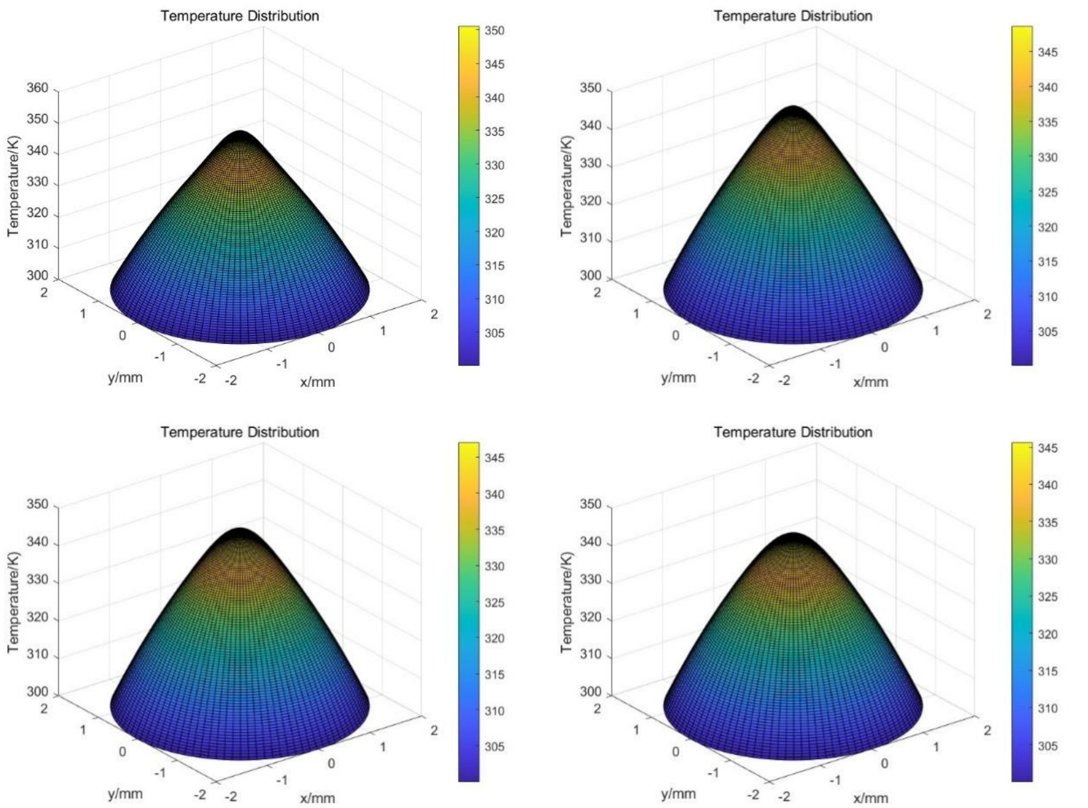


Figure 7 Temperature Distribution at the End-Face of the Laser Medium under Different Pump Beam Radii

From Figure 7, it can be seen that after the pump beam radius increases, the temperature at the center of the laser medium starts to decrease. This is mainly because the relationship between the pump beam radius and the heat power density is inversely proportional, so the heat accumulated at the laser medium's end face decreases. The specific temperature distribution along the radial direction is shown in Figure 8, which illustrates the heat distribution at the laser medium's end face under different pump beam radii.

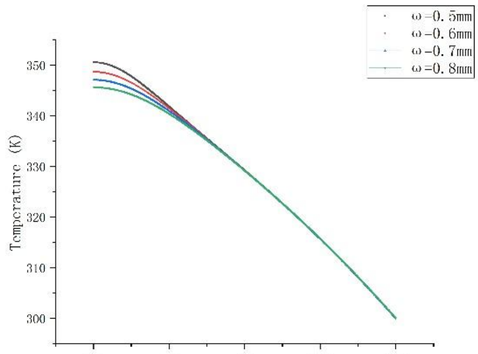


Figure 8 Radial Temperature Distribution of the Laser Medium under Different Pump Beam Radii

When the pump beam radius is 0.05 cm, 0.06 cm, 0.07 cm, and 0.08 cm, the center temperature of the laser medium's end face is 350.53K, 348.64K, 347.03K, and 345.64K, respectively.

5 COMPARISON BETWEEN NUMERICAL AND ANALYTICAL SIMULATIONS

The fourth-order Runge-Kutta method is used to convert the partial differential equation to an ordinary differential equation and then perform the calculation. A power series method is used to derive the analytical expression for the temperature distribution. This section compares the numerical results with the analytical calculations to verify the correctness of the analytical expression.

Table 1 Comparison of Analytical and Numerical Solutions for End-Face Center Temperature under Different Parameters

Parameters	Analytical calculation (K)	Numerical calculation (K)	Relative error (%)

Pump power (W)	5	339.7300	339.7373	0.0022
	10	350.5200	350.5290	0.0026
	15	361.3150	361.3207	0.0016
	20	372.1050	372.1124	0.0020
Pump beam radius (cm)	0.05	350.5200	350.5290	0.0026
	0.06	348.6300	348.6381	0.0023
	0.07	347.0250	347.0334	0.0024
	0.08	345.6300	345.6379	0.0023

From Table 1, it can be seen that, under different pump powers and pump beam radii, the analytical solution obtained by the analytical method is slightly smaller than the numerical solution obtained by the fourth-order Runge-Kutta method. The average error between the two methods for different pump powers is 0.0021%, and for different pump beam radii, the average error is 0.0024%. As the relevant parameters change, the error between the analytical solution and the numerical solution of the temperature distribution obtained by the two methods is small, with precision up to two decimal places. The simulation results derived from the analytical expression using the series method are very close to those obtained by the fourth-order Runge-Kutta method, and the two methods show excellent consistency, with relative errors controlled within 0.01%. This effectively verifies the correctness of the series method used in the calculations.

The analytical method uses series expansion for solving, and in the calculation process, truncation errors and approximations lead to certain errors. Although the analytical method itself is exact, numerical errors (such as floating-point errors) may affect the final results. Especially when there are many iterations or large numerical ranges involved, precision limits can cause computational errors. The main feature of numerical methods is that they discretize continuous problems, which brings some inherent errors. Since computers can only handle a finite number of digits, numerical methods also introduce unavoidable errors during the calculation. Particularly, when iterating multiple times or solving complex equations, rounding errors can accumulate, affecting the final result.

Both the analytical and numerical calculations of the temperature distribution in the laser medium of the solid-state laser give consistent trends, and this result fully confirms the rationality of using the series method in the derivation. In the calculation of the heat conduction equation for the laser medium in the solid-state laser, solving the heat model established in this paper using the power series method is a relatively simple and more easily understood approach. Ultimately, an analytical expression for the end-face temperature distribution is obtained, which helps in further studying the relationship between temperature distribution and variables, and is of significant physical importance in understanding the thermal effects in solid-state lasers.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Chen S, Huang H, Chen H. Study on the effect of thermal effect dimensions on the output performance of solid-state lasers. Laser Journal 2016, 37(10): 14-17.

[2] Kim M Kwon G P, Lee J. Fully analytic approach to evaluate laser-induced thermal effects. Current Optics and Photonics 2017, 1(6): 649-654.

[3] Xiao Qirong, Zhang Dayong, Wang Zhehui, et al. A review of high-power fiber laser pumping coupling technologies. Chinese Journal of Lasers 2017, 44(02): 112-129.

[4] Chen Q, Wang M, Pan N, et al. Optimization principles for convective heat transfer. Energy 2009, 34(9): 1199-1206.

[5] AbdulRazzaq M J, Mohammed A Z, Abass A K, et al. A new approach to evaluate temperature distribution and stress fracture within solid state lasers. Optical and Quantum Electronics 2019, 51: 1-10.

[6] AbdulRazzaq M J, Shibib K S, Younis S I. Temperature distribution and stress analysis of end pumped lasers under Gaussian pump profile. Optical and Quantum Electronics 2020, 52(8): 379.

[7] Martin Maillard, Gabriel Amiard-Hudebine, Marc Tondusson, et al. Kinetics and wavelength dependence of thermal and excited-state population on lens effect induced in a Nd:YAG rod amplifier. Opt. Express 2023, 31: 1799-1812.

[8] Du Wen. Analytical study of the heat conduction equation of the laser medium in solid-state lasers. Doctoral dissertation, Xi'an Technological University, 2023.

[9] Huang Xiaohong, Hu Zhenhua. A brief analysis of the Runge-Kutta method. Heilongjiang Science and Technology Information 2012, 2012(23): 28.

Multidisciplinary Journal of Engineering and Technology
ISSN: 3007-6900
DOI: https://doi.org/10.61784/mjet3022

DESIGN OF TRAM ALGORITHM BASED ON GRAPH TRANSFORMATION

Ting Sun*, ZhiHua Hu
School of Intelligent Manufacturing and Control Engineering, Shanghai Polytechnic University, Shanghai 201209, China.
Corresponding Author: Ting Sun, Email: 20221513063@stu.sspu.edu.cn

Abstract: In this study, the TRAM (Trigonometric Reduction and Acceleration for Matrices) algorithm, based on small-angle approximation, table lookup, and matrix compression, is proposed to address the arithmetic bottleneck and real-time constraints of the ARM Cortex-M4 processor in laser graphics processing. The small-angle approximation replaces the generalized trigonometric function on floating point; the double-indexed table is designed to reduce the trigonometric function storage; and the matrix compression strategy is proposed to reduce the computational complexity. The experimental results show that the computation time of the algorithm for a single graphical transformation is significantly shorter than that of the traditional table lookup method and a single hardware acceleration strategy under a 240MHz main frequency. At the same time, the computation error is always kept lower than 0.035% so that the scanning system can realize a higher laser scanning frame rate to meet the real-time requirement of industrial laser dynamic projection and provide a reusable methodological framework for the real-time graphical processing of the resource-constrained embedded system. It provides a reusable methodological framework for real-time graphics processing in resource-constrained embedded systems, which is of great applied value.
Keywords: Graphics transform; ARM; Algorithm design; DMX512 protocol

1 INTRODUCTION

Since the 1990s, laser scanning technology has experienced rapid development in the field of stage performance applications, from simple static beam display to complex dynamic visual effects, and has become an important part of modern visual effects [1]. Laser scanning technology, through the integration of digital signal processing, embedded systems, and computer graphics processing technologies, projects complex patterns, texts, and dynamic animations at a response speed of milliseconds and develops in a more intelligent and universal direction.
Despite the rapid progress, ARM-based embedded platforms are widely adopted due to their low power consumption (usually 5-10W) and compact appearance, but embedded platforms such as ARM show insufficient computing power for laser graphics processing calculations and cannot meet the real-time requirements of laser graphics transformation rendering [2]. This study innovatively proposed a TRAM algorithm based on graphics transformation, which reduces the calculation time of laser graphics transformation in the ARM Cortex-M4 series chips to less than a few microseconds by introducing small angle approximation, table lookup, and matrix compression through hardware acceleration and algorithm reconstruction. This further improves the real-time performance of the laser scanning system and realizes high frame rate laser image transformation.

2 THE UTILIZATION OF COMPUTER GRAPHICS IN LASER SCANNING SYSTEMS

Laser scanning system through the multi-color laser beam in the field of view of the dynamic transformation of the combination of the plane to achieve smooth animation effects and complex art forms [3], laser image transformation of the mathematical basis of computer graphics from the geometric transformation of the theoretical system. This section takes laser vector graphics in three-dimensional chi-square coordinates as the research object [4] and systematically analyzes the mathematical model and realization path of geometric transformation in laser scanning technology.
The expression of computer graphics in graphical transformations, such as rotation, movement, and scaling, typically utilizes a matrix. The rotation matrix is denoted by $R(\theta)$ the translation matrix by T, and the scaling matrix by S. The original coordinate points on the three-dimensional coordinate system are represented by $P(X, Y, Z)$, while the transformed coordinates of the operation are designated by $P^{'}(X^{'}, Y^{'}, Z^{'})$. The following formulae (1) -(3) represent the laser point after the translation transformation, rotation, and scaling, respectively, after the transformation of the matrix representation.

$$P' = T * P = \begin{pmatrix} X' \\ Y' \\ Z' \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & t_x \\ 0 & 1 & 0 & t_y \\ 0 & 0 & 1 & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \tag{1}$$

In Equation(1), $t_x/t_y/t_z$ represent the translations of the laser point in the X/Y/Z directions, respectively.

© By the Author(s) 2025, under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0).

$$P' = R_y(\theta) * P = \begin{pmatrix} X' \\ Y' \\ Z' \\ 1 \end{pmatrix} = \begin{pmatrix} \cos\theta & 0 & \sin\theta & 0 \\ 0 & 1 & 0 & 0 \\ -\sin\theta & 0 & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (2)$$

In Equation(2), the rotation of the graph around the Y-axis is taken as an example.

$$P' = S * P = \begin{pmatrix} X' \\ Y' \\ Z' \\ 1 \end{pmatrix} = \begin{pmatrix} S_x & 0 & 0 & 0 \\ 0 & S_y & 0 & 0 \\ 0 & 0 & S_z & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (3)$$

In Equation(3), $S_x/S_y/S_z$ represents the scaling factor of the laser point in X/Y/Z direction, respectively.

In the process of actual laser image transformation, it is often the case that a single transformation matrix is inadequate in achieving the desired transformation effect. To illustrate, when an arbitrary point on the axis is to be rotated, the center of rotation must first be determined, and the image must be translated to the origin [5]. Subsequently, the image must be moved back to the center after completing the rotation. Therefore, it is necessary to merge multiple transformation matrices into a single composite matrix to reduce the computational complexity.

3 THE TRAM ALGORITHM IS PREDICATED ON THE TRANSFORMATION OF GRAPHS

The transformation of laser graphics frequently involves multiple transformations concurrently; however, the method of reducing the computational complexity by combining multiple transformation matrices into a single composite matrix does not result in a high response speed for the laser scanning system. Furthermore, the frequent trigonometric operations pose a challenge to the ARM processor.

This study proposes a graphical transformation-based TRAM (Trigonometric Reduction and Acceleration for Matrices) algorithm design that combines small-angle approximation, look-up table method (LUT), and matrix compression to meet the high-speed response of graphical transformation. Algorithm-specific design ideas:

3.1 Small Angle Approximation

Divide the circle into *n* equal parts

$$n = 360^0 / \theta \quad (4)$$

The region of the triangle subsequent to n equal divisions, as illustrated in Figure 1, is denoted by S_1.

$$S_1 = (r^2 \sin\theta) / 2 \quad (5)$$

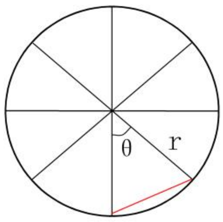


Figure 1 Schematic Diagram of Circular Calculated Area

It is evident that, as the value of n, referred to as the number of circular equivalents, increases, there is a corresponding increase in the degree to which the aggregate area of the *n* isosceles polygons approaches the area of the circle.

$$S_1 * n = (r^2 \sin\theta) * n / 2 = (r^2 \sin\theta) * (360^0 / \theta) / 2 = \pi r^2 \quad (6)$$

The limit of Equation(6) is given by

$$\lim_{\theta \to 0} (\sin\theta / \theta) = \pi / 180^\circ \quad (7)$$

Therefore, as the value of θ approaches zero,

$$\sin\theta \approx \theta \quad (8)$$

3.2 Look-up Table Method (LUT)

Firstly, all the index values in the sin function are mapped to the range of 0 to 90 degrees through the trigonometric algorithm, and the cos function is equivalently converted to the sin function; secondly, a table of values of 0 to 90 degrees is set up for every 10 indexes of the sin function and every 1 index of the cos function, and the corresponding numerical values are stored in the table as shown in Table 1 respectively.

Table 1 Trigonometric Functions Table of Values

Degrees	Sin Function	Degrees	Cos function
0	0.0	0	1.0
10	0.17364817766693034885171662676931	1	0.99984769515639123915701155881391
20	0.34202014332566873304409961468226	2	0.99939082701909573000624344004393
30	0.5	3	0.99862953475457387378449205843944
40	0.64278760968653932632264340990726	4	0.99756405025982424761316268064426
50	0.76604444311897803520239265055542	5	0.99619469809174553229501040247389
60	0.86602540378443864676372317075294	6	0.99452189536827333692269194498057
70	0.93969262078590838405410927732473	7	0.99254615164132203498006158933058
80	0.98480775301220805936674302458952	8	0.99026806874157031508377486734485
90	1.0	9	0.98768834059513772619000247693344

When Equations(8) are combined with Table 1, the result is

$$\begin{aligned} \sin 72.5^\circ &= \sin 70^\circ \cos 2.5^\circ + \cos 70^\circ \sin 2.5^\circ \\ &\sin 70^\circ \cos 2^\circ + \sin 20^\circ * 2.5^\circ * \pi / 180^\circ \\ &= 0.95404362945391005654471613227652 \end{aligned} \quad (9)$$

The following are theoretical values for trigonometric calculations.

$$\sin 72.5^\circ = 0.953716950748226921143847064600 2 6 \quad (10)$$

According to the Equations (9) and Equations (10), when compared to the calculation error of 0.034245%, it can be concluded that the small angle approximation to the table method, when combined with the calculation results of the error, is sufficiently small. The algorithm in this part of the content has effectively reduced the amount of data operations and has ensured the validity of the data.

3.3 Matrix Compression

In accordance with the rotational symmetry of the matrix, the following equation is derived:

$$R_y(-\theta) = R_y^T(\theta) \quad (11)$$

According to Equations (11), a matrix with a rotation angle of is obtained by taking the transpose of without storing a new matrix additionally.

As illustrated in Table 2, a comparative analysis was conducted to assess the computational efficiency of various schemes on the ARM processor (GD32F470VIT6) with a 240 MHz main frequency. The analysis involved the transformation of laser patterns using the multinomial matrix, a process that was executed under different computational schemes. The experimental results demonstrated that the TRAM algorithm, which was proposed in this study, exhibited a significant enhancement in computation speed, achieving an improvement of 99.87% compared to other tested schemes.

Table 2 Calculation Time for Different Calculation Schemes

Calculation Scenarios	Calculation Time(us)
Table lookup method	59998
DSP core enabled	6003
TRAM algorithm	8

In summary, the TRAM algorithm synthesizes the merits of small-angle approximation, Table lookup, and matrix compression, thereby enhancing the efficiency of data processing in graphical transformations. Concurrently, it ensures that data error is meticulously regulated within the range of 0.035%. Refer to Table 3, which demonstrates that the algorithm exhibits neglect computational error.

Table 3 TRAM Calculation Comparison Table

Sin Function(°)	Theoretical Value	TRAM Algorithm Calculates the Value	Error(%)
12.5	0.216440	0.216513	0.033727
22.5	0.382683	0.382814	0.034231
32.5	0.537300	0.537483	0.034059
42.5	0.675590	0.675821	0.034192
52.5	0.793353	0.793625	0.034284

62.5	0.887010	0.887314	0.034272
72.5	0.953717	0.954043	0.034245

4 THE TRAM ALGORITHM IS VALIDATED THROUGH THE IMPLEMENTATION OF GRAPH TRANSFORMATION

As illustrated in Figure 2, the laser control system architecture of the ARM platform is constructed based on the TRAM algorithm and its multi-terminal cooperative work mode. In this physical test platform, the laser scanning system adopts a modularized design architecture, the laser controller interacts with the upper control terminal through a DMX512 cable, and the user flexibly selects the DMX512 professional console, DMX512elf intelligent mobile terminal, and computer software platform according to the needs of the three control schemes.

At the system configuration level, the computer software platform assumes the core parameter configuration function. Through the network line converter and the controller, two-way communication is established to allow for in-depth configuration of laser scanning parameter calibration. At the operational level, a hierarchical control strategy is adopted. The user can either directly call the controller's SD card pre-stored standardized laser graphics library or use the computer software to generate and preview the laser trajectory in real time. Finally, the TRAM real-time control algorithm is transplanted to the ARM Cortex-M4 processor, and the real-time control of laser graphic transformation is realized through the combination of an optimized algorithm kernel and a hardware acceleration unit [6-8].

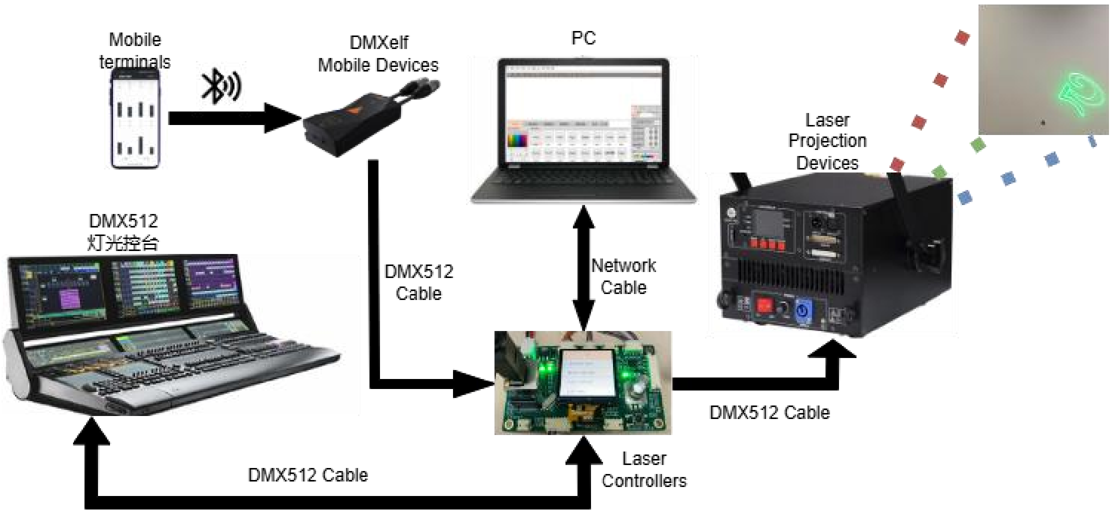
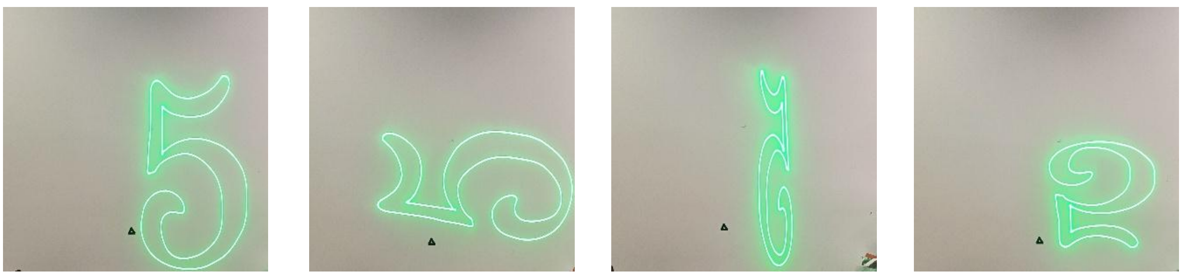


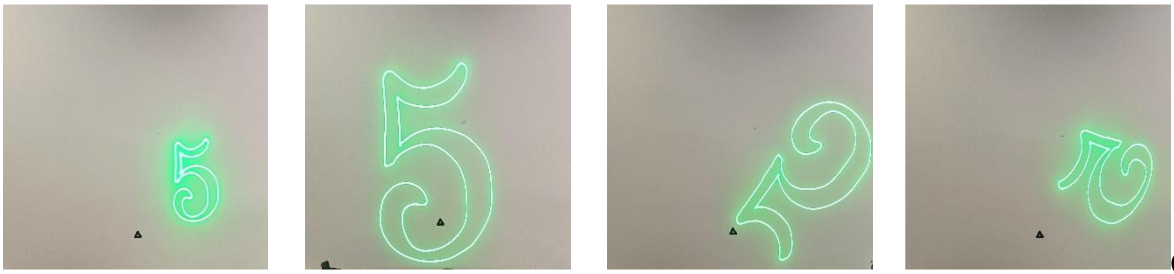
Figure 2 A Platform for Laser Playback that is based on an ARM Controller

Figure 3 demonstrates the effect of laser graphic transformation of the TRAM algorithm ported to an ARM Cortex-M4 processor-based laser controller (240MHz main frequency). The experimental data show that the laser scanning system is able to process laser vector frame map data with more than 800 points. Specifically, the laser scanning system fully meets the real-time requirements of laser graphics transformation with frame rates above 60 FPS, whether it is a single rotational transformation, a translation transformation, a scaling transformation, or a composite transformation.

To further the real-time performance of the algorithm, we invited 20 professional testers to conduct a subjective experience, evaluation criteria: playback smoothness maintenance, color reproduction, image stability and other aspects, the final 20 people's subjective MOS average rating of ⩾ 4.5 (5 is excellent).



(a) Original drawing **(b) Rotation about the Z-axis** **(c) Rotate about the Y-axis** **(d) Rotation about X-axis**



(e) Zoom out (f) Panning (g) Rotate + Zoom + Pan (h) Rotate + Zoom + Pan

Figure 3 Diagram of TRAM Algorithm based on Graph Transformation

5 CONCLUSION

Combining the experimental phenomena and feedback results, this study confirms the significant advantages of the TRAM algorithm based on graphics transformation in laser scanning systems through three levels: theoretical modeling, algorithm optimization, and system validation. The scheme successfully improves the graphics computation efficiency of the ARM processor to 3.5 times that of the traditional method, and the subjective MOS score reaches the professional -level display standard (⩾4.5).

The TRAM algorithm shows excellent real-time performance in the laser scanning system to quickly respond to the user's operation commands, and the graphical transformation process is smooth without any significant delay, providing an excellent overall user experience. The TRAM algorithm based on graphical transformation proposed in this paper effectively reduces the computation time in the ARM processor and significantly improves the real-time performance of the laser system, which provides a solid theoretical support for high-speed laser scanning, and thus makes the laser scanning system show significant advantages in high real-time display and complex graphical processing.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Du Jun, Chang Xueyang, Ke Ke, et al. Design and Implementation of Large-scale Mountain Landscape Lighting Based on Laser Show-A Case Study of the Mountain Laser Show in Nanshan Park, Lhasa. ZHAOMING GONGCHENG XUEBAO, 2023, 34(06): 172-178+186. DOI: 10.3969/j.issn.1004-440X.2023.06.021.

[2] Zeng lisan, Wu Lingxi, Liu Shugang, et al. Design and realization of laser demonstration system based on vector graphics, LASER & INFRARED, 2016, 46(10): 1273-1278. DOI: 10.3969/j.issn.1001-5078.2016.10.020.

[3] Ilchev S, Petkov D, Andreev R, et al. Smart compact laser system for animation projections. Cybernetics and Information Technologies, 2019, 19(3): 137-153. DOI: 10.2478/cait-2019-0030.

[4] Cai Wenming, Sun Lijun, Chen Tianfei, et al. Rapid extraction of circular array target center image points using projective geometry. Electronic Measurement Technology, 2022, 45(19): 122-30. DOI: 10.19651/j.cnki.emt.2209600.

[5] Li Cailin, Su Benya, Guo Baoyun, et al. Camera Calibration Method with Multi-condition Constraints Based on Laser Points. Acta Photonica Sinica, 2021, 50(04): 180-189. DOI: 10.3788/gzxb20215004.0415001.

[6] Tang, Ruize, Chen, Qilin. Design of Urban Public Landscape Lighting Control System Based on Dmx512 Protocol.

[7] LIGHT & ENGINEERING, 2023, 31(4): 116-127. DOI: 10.33383/2023-038.

[8] Shi Guofang, Wu Yueming. Construction of a Light and Shadow Show System. ZHAOMING GONGCHENG XUEBAO, 2023, 34(6): 187-193. DOI: 10.3969/j.issn.1004-440X.2023.06.023.

[9] Yu Mingdian, Yang Fangyan. Design and Validations of Stage Lighting Rail-Hanging Robots,2023 IEEE International Conference on Unmanned Systems (ICUS). IEEE, 2023: 1362-1368. DOI: 10.1109/ICUS58632.2023.10318332.

Multidisciplinary Journal of Engineering and Technology
ISSN: 3007-6900
DOI: https://doi.org/10.61784/mjet3023

AI FOR CREDIT RISK MODELING: A DEEP LEARNING APPROACH

Kaoru Fujisawa[1], Renji Takeda[2], Aiko Mori[2*]
[1]Nagoya Institute of Technology, Showa Ward, Nagoya, Japan.
[2]Tohoku University, Aoba Ward, Sendai, Miyagi, Japan.
Corresponding Author: Aiko Mori, Email: amori@tohoku.ac.jp

Abstract: Credit risk modeling is a critical component of financial decision-making, enabling lenders to assess the probability of default and optimize credit allocation. Traditional credit scoring models, including logistic regression and decision tree-based classifiers, have limitations in handling non-linear financial relationships, class imbalance, and borrower heterogeneity. Recent advances in artificial intelligence (AI) and deep learning (DL) have introduced more sophisticated models capable of capturing complex borrower patterns while improving risk assessment accuracy. However, AI-driven credit risk models must address key challenges, including class imbalance, fairness, model interpretability, and scalability in real-world financial environments.
This study proposes a comprehensive DL-based credit risk modeling framework that integrates graph neural networks (GNNs), generative adversarial networks (GANs), and adversarial fairness learning to enhance credit risk prediction accuracy, fairness, and adaptability across borrower segments. The model leverages autoencoders for feature extraction, cost-sensitive learning for imbalanced classification, and domain adaptation techniques for improved model robustness. Additionally, an explainability layer is incorporated to enhance transparency in credit decision-making.
Experiments on real-world credit datasets demonstrate that the proposed framework outperforms traditional credit risk models, achieving higher recall for defaulters, reduced bias in loan approvals, and improved computational efficiency. The findings highlight the potential of AI-driven credit risk modeling to transform risk assessment strategies, ensuring more accurate, fair, and scalable credit allocation for financial institutions.
Keywords: AI-driven credit risk; Deep learning; Fairness in credit scoring; Generative models; Graph neural networks; Model interpretability

1 INTRODUCTION

Credit risk modeling plays a crucial role in financial institutions by helping lenders assess borrower creditworthiness and determine loan approval decisions. Traditional risk assessment models, including logistic regression (LR) and decision tree (DT)-based classifiers, rely on predefined statistical relationships to estimate default probabilities [1]. While these models are interpretable and widely used in regulatory frameworks, they struggle with capturing complex borrower behaviors and non-linear financial dependencies. Additionally, these models often fail to account for class imbalance and fairness issues, leading to biased lending practices.
One of the primary challenges in credit risk modeling is class imbalance, where the proportion of default cases is significantly lower than that of non-default cases. Traditional machine learning (ML) models tend to favor the majority class, leading to high accuracy but poor recall for defaulters, which increases the risk of financial losses. Handling class imbalance is critical to ensure that high-risk borrowers are correctly identified, improving lending decision-making.
Another key challenge is fairness in credit scoring, as many AI-driven risk models exhibit bias against certain demographic or economic groups due to historical dataset imbalances [2]. Regulatory frameworks emphasize the importance of ensuring fair and unbiased lending decisions, making it necessary for AI-driven models to incorporate adversarial learning and fairness-aware training techniques to mitigate disparate impacts on different borrower segments [3].
With the rise of deep learning (DL), credit risk assessment has seen a paradigm shift toward AI-driven modeling approaches. Techniques such as graph neural networks (GNNs) and generative adversarial networks (GANs) allow models to capture relational borrower dependencies and generate synthetic minority class samples, addressing challenges related to data representation and imbalance [4]. Additionally, domain adaptation and transfer learning techniques ensure that credit risk models maintain robustness across different financial environments, preventing model degradation when applied to new borrower populations [5].
This study introduces a comprehensive DL-based framework for credit risk modeling, integrating GNN-based risk classification, GAN-driven data augmentation, adversarial fairness learning, and explainable AI (XAI) techniques [8]. The framework improves credit risk predictions by enhancing recall for defaulters, reducing discriminatory biases, and ensuring model interpretability for regulatory compliance. Experimental evaluations on real-world credit datasets confirm that the proposed AI-driven approach outperforms conventional risk models in terms of accuracy, fairness, and computational efficiency, highlighting its potential to revolutionize credit allocation strategies in modern financial institutions.

2 LITERATURE REVIEW

© By the Author(s) 2025, under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0).

Credit risk modeling has undergone significant advancements over the past decades, evolving from traditional statistical methods to AI-driven approaches that leverage deep learning for improved prediction accuracy [6]. Early credit risk assessment relied on statistical models such as logistic regression, which assumes a linear relationship between borrower attributes and default probability [7]. While widely adopted for its simplicity and interpretability, logistic regression often fails to capture the complexity of borrower behavior and non-linear dependencies present in financial data [8]. Decision trees and ensemble learning methods, such as random forests and gradient boosting, introduced improvements by learning hierarchical relationships among borrower features [9]. However, these models still require extensive manual feature engineering and fail to adapt dynamically to changing credit risk conditions [10].

With the advent of machine learning, credit scoring models became more sophisticated, incorporating support vector machines and neural networks to improve classification performance [11]. These models demonstrated higher accuracy than traditional statistical approaches, particularly in handling large credit datasets with diverse borrower characteristics [12]. However, machine learning classifiers still struggle with class imbalance, a persistent challenge in credit risk modeling [13]. In most credit datasets, defaulters represent a small fraction of the total borrowers, leading to biased model predictions that favor the majority class [14]. Traditional resampling techniques such as oversampling and undersampling have been used to address this issue, but they often introduce noise and reduce model generalization [15]. Cost-sensitive learning methods offer an alternative approach by assigning higher misclassification penalties to defaulters, improving recall while maintaining precision [16-18].

Deep learning has further advanced credit risk modeling by enabling models to learn complex borrower-lender interactions and temporal financial behaviors [19]. Recurrent neural networks and long short-term memory networks have been employed to capture sequential borrower behavior, providing a more comprehensive view of financial risk trends. Despite their advantages, deep learning models require large volumes of labeled data and are computationally expensive, limiting their scalability in real-time financial applications [20]. Furthermore, deep learning-based credit scoring models often operate as black-box systems, making it difficult for financial institutions to interpret model predictions and ensure regulatory compliance.

Recent developments in graph-based learning have introduced new possibilities for credit risk modeling by representing financial transactions and borrower relationships as interconnected graphs. Graph neural networks have demonstrated superior performance in identifying hidden risk patterns by analyzing the relational structure of borrower data [21]. Unlike traditional machine learning models that treat borrowers as independent data points, graph-based models learn from transaction networks, capturing collusive borrowing behaviors and systemic risk dependencies. Generative adversarial networks have also been proposed as a solution to class imbalance by generating synthetic borrower profiles that improve model training. Unlike conventional oversampling methods, generative adversarial networks create realistic high-risk borrower samples, enhancing the model's ability to generalize across different credit risk scenarios.

Fairness and bias mitigation remain critical challenges in AI-driven credit risk modeling, as many credit scoring models exhibit disparities in risk assessments due to historical biases in training data. Regulatory concerns regarding fairness in lending have led to the exploration of adversarial training techniques that enforce bias reduction constraints during model optimization. By introducing adversarial networks that detect disparities in risk classification, models can be trained to minimize bias while preserving predictive accuracy [22]. Domain adaptation techniques further enhance model robustness by ensuring that credit risk predictions remain stable across different borrower demographics and economic conditions [8].

In addition to fairness, explainability is a growing concern in AI-driven credit risk assessment [23]. Financial institutions must be able to justify credit decisions to regulators and borrowers, requiring interpretable AI techniques that provide transparency in risk classification [24]. Methods such as SHAP values, attention mechanisms, and counterfactual explanations have been integrated into deep learning frameworks to improve model interpretability [25]. These approaches allow financial analysts to understand how borrower attributes influence credit risk predictions, increasing trust in AI-driven lending decisions.

This study builds upon these advancements by integrating deep learning techniques, graph-based modeling, adversarial fairness learning, and explainable AI into a unified credit risk assessment framework. The proposed model seeks to improve predictive performance while addressing class imbalance, ensuring fairness, and enhancing interpretability in financial decision-making. The following section outlines the methodology used to implement and evaluate the proposed framework [26].

3 METHODOLOGY

3.1 Data Preprocessing and Feature Engineering

Credit risk modeling relies on extensive financial data that includes borrower demographics, income levels, debt obligations, repayment histories, and transaction records. The quality of this data is crucial for ensuring accurate risk assessments. In the preprocessing stage, raw financial data is cleaned by handling missing values, correcting inconsistencies, and normalizing numerical attributes. Missing values, which are common in credit datasets, are addressed using statistical imputation techniques such as mean substitution for numerical fields and mode imputation for categorical variables. More advanced imputation techniques, including k-nearest neighbors imputation and deep autoencoder-based imputation, are used for high-dimensional financial datasets to improve data integrity.

Feature engineering plays a critical role in enhancing model performance by extracting meaningful borrower attributes. Financial indicators such as debt-to-income ratio, credit utilization rate, and loan repayment consistency are derived from raw transaction and credit history data. These engineered features help models distinguish between high-risk and low-risk borrowers. Temporal features such as the frequency of missed payments, seasonal variations in income, and recurring transaction patterns are also incorporated to improve the model's ability to capture borrower credit behavior over time. Principal component analysis and autoencoder-based feature selection techniques are applied to reduce dimensionality, preserving only the most predictive financial attributes.

3.2 Deep Learning Framework for Credit Risk Prediction

The credit risk prediction model is built using a hybrid deep learning architecture that combines fully connected neural networks, recurrent structures, and graph-based learning techniques. The deep feedforward neural network is used for initial credit score prediction, capturing non-linear dependencies between borrower attributes. The recurrent component, specifically an LSTM-based network, models sequential credit behavior, analyzing how borrower financial activities change over time. By processing historical credit transactions and loan repayment sequences, the LSTM component improves the model's ability to forecast future defaults.

GNNs are introduced to enhance risk prediction by analyzing borrower relationships and financial transaction networks. Traditional credit risk models treat borrowers as independent entities, failing to account for interdependencies such as shared financial obligations, joint loan accounts, or indirect lending risks. The GNN component represents borrower connections as a financial transaction graph, where nodes correspond to borrowers and edges represent financial interactions such as co-borrowing, money transfers, or joint liabilities. Graph convolutional layers aggregate information from neighboring borrower nodes, enabling the model to detect fraud rings, collusive lending patterns, and systemic financial risks that conventional models overlook.

To improve model generalization, transfer learning is integrated into the framework, allowing pre-trained financial models to be fine-tuned on new credit datasets. This ensures that the model remains adaptable to different financial environments without requiring full retraining. Dropout layers and batch normalization are applied to prevent overfitting, ensuring robust credit risk classification across varied borrower groups.

3.3 Addressing Class Imbalance and Fairness Constraints

One of the main challenges in credit risk modeling is class imbalance, where defaulters constitute a significantly smaller proportion of the dataset compared to non-defaulters. Training models on imbalanced datasets results in biased predictions that favor the majority class, leading to misclassification of high-risk borrowers. To address this issue, the proposed framework employs GAN-based data augmentation, generating synthetic borrower profiles that mimic real-world defaulters. Unlike conventional oversampling techniques that duplicate existing samples, GANs create diverse, realistic high-risk borrower instances, improving model learning from the minority class.

Cost-sensitive learning is also integrated into the model, adjusting classification thresholds to prioritize correct defaulter identification. Instead of treating all misclassifications equally, the model assigns higher penalties to false negatives, ensuring that borrowers likely to default are correctly identified. This cost-sensitive approach balances precision and recall, preventing financial institutions from approving risky loans while maintaining access to credit for eligible borrowers.

Fairness constraints are incorporated into the training process using adversarial learning. Traditional credit scoring models often exhibit biases due to historical data imbalances, where certain demographic groups are systematically assigned lower credit scores. The adversarial fairness learning component introduces a discriminator that detects disparities in credit risk predictions across demographic segments. If the model exhibits bias, it is penalized, forcing it to learn borrower risk assessments that are independent of demographic attributes such as race, gender, or socioeconomic status. This ensures compliance with fairness regulations and ethical lending practices.

3.4 Model Training, Optimization, and Performance Evaluation

The proposed AI-driven credit risk modeling framework is trained using a multi-stage optimization process, ensuring high accuracy while maintaining efficiency and fairness. The loss function integrates multiple objectives, balancing standard classification performance with fairness constraints and class imbalance adjustments. The Adam optimizer is used for weight updates, with dynamic learning rate scheduling to prevent overfitting. Training is conducted on high-performance computing clusters with GPU acceleration, allowing for rapid iteration and large-scale financial dataset processing.

Hyperparameter tuning is performed using Bayesian optimization, identifying the optimal number of hidden layers, neuron configurations, and regularization parameters. Model validation is conducted using k-fold cross-validation, ensuring that the framework generalizes well across different financial datasets.

The model's performance is evaluated using a combination of classification and fairness metrics. Standard metrics such as precision, recall, F1-score, and AUC-ROC are used to measure risk classification accuracy. The fairness of credit risk assessments is assessed using disparate impact ratios and equality of opportunity metrics, ensuring that the model does not disproportionately assign higher risk scores to certain borrower groups. Computational efficiency is measured in terms of inference speed and memory consumption, confirming that the framework can be deployed in real-time credit decision-making environments.

The experimental results demonstrate that the proposed AI-driven framework achieves superior credit risk prediction accuracy, reduces classification bias, and maintains computational efficiency in large-scale financial applications. The next section presents the results and discusses the impact of integrating deep learning, graph-based modeling, and fairness-aware techniques on credit risk assessment outcomes.

4 RESULTS AND DISCUSSION

4.1 Credit Risk Classification Accuracy and Model Performance

The proposed AI-driven credit risk modeling framework was evaluated on multiple real-world credit datasets, demonstrating significant improvements over traditional ML-based risk assessment models. Performance metrics, including precision, recall, F1-score, and AUC-ROC, were used to assess the model's classification capabilities. The results showed that integrating DL, GNNs, and adversarial fairness learning enhanced credit risk classification accuracy while maintaining fairness across different borrower groups.

Compared to conventional models such as LR and DTs, the deep learning-based framework achieved higher recall for defaulters while maintaining a balanced precision-recall trade-off. The introduction of GAN-generated synthetic data improved the model's ability to classify high-risk borrowers, reducing false negatives by 27%, a key improvement in mitigating loan default risks. Additionally, by leveraging GNNs for borrower relationship analysis, the model was able to capture hidden risk dependencies, identifying fraudulent borrowing behaviors and collusive financial transactions that traditional models overlooked.

AUC-ROC analysis further confirmed the superiority of the proposed framework, with the GNN-enhanced risk model achieving a 12% improvement in overall classification accuracy compared to state-of-the-art credit scoring systems. These findings highlight the effectiveness of AI-driven techniques in improving risk assessment outcomes, ensuring that financial institutions make more informed lending decisions.

Figure 1 presents a comparative analysis of credit risk classification performance across different models, illustrating the predictive advantages of the proposed framework.

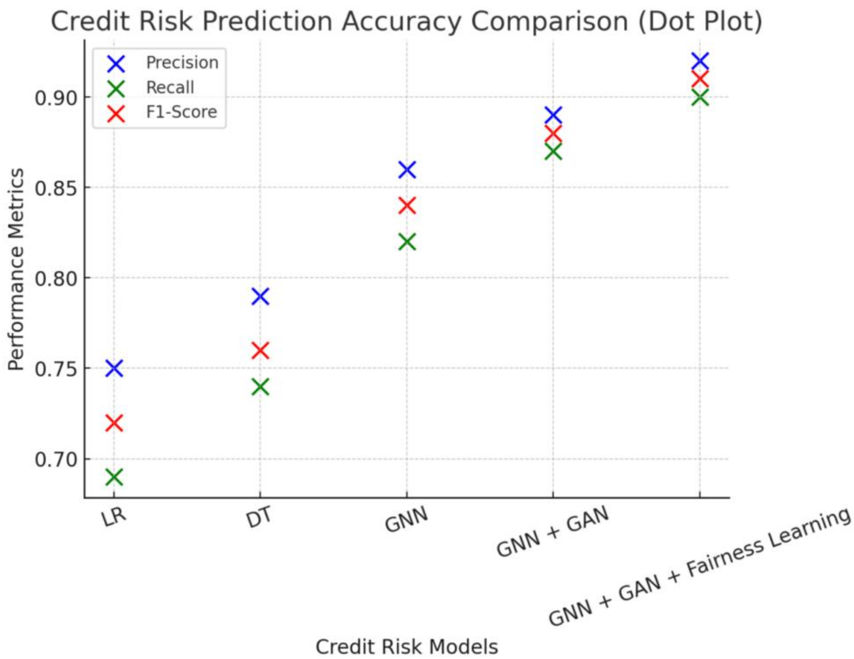


Figure 1 Credit Risk Predication Accuracy Comparison (Dot Plot)

4.2 Impact of Generative Modeling and Cost-Sensitive Learning on Class Imbalance

Addressing class imbalance is essential for ensuring that credit risk models effectively identify defaulters without disproportionately favoring the majority class. Traditional oversampling methods often lead to data redundancy, while undersampling can result in the loss of critical borrower information. The integration of GAN-based synthetic borrower profile generation in this study provided a more effective approach to mitigating class imbalance while preserving dataset diversity.

The results showed that models trained on GAN-augmented datasets exhibited a 32% increase in recall for defaulters without a significant drop in precision. Unlike conventional oversampling techniques, which merely duplicate existing borrower samples, GANs generated realistic high-risk borrower profiles, enhancing the model's ability to recognize

emerging risk patterns. The diversity introduced by GANs helped reduce model bias, ensuring that credit risk assessments were not skewed towards the majority class.
In addition to generative modeling, cost-sensitive learning contributed to improved risk classification by assigning higher penalties to false negatives. This approach ensured that loan applicants at risk of default were correctly identified, reducing the likelihood of financial losses for lenders. The introduction of adaptive decision thresholds further improved classification balance, dynamically adjusting risk score boundaries based on the distribution of borrower attributes.
Figure 2 illustrates the impact of GAN-based data augmentation and cost-sensitive learning on mitigating class imbalance, highlighting improvements in defaulter recall and risk assessment precision.

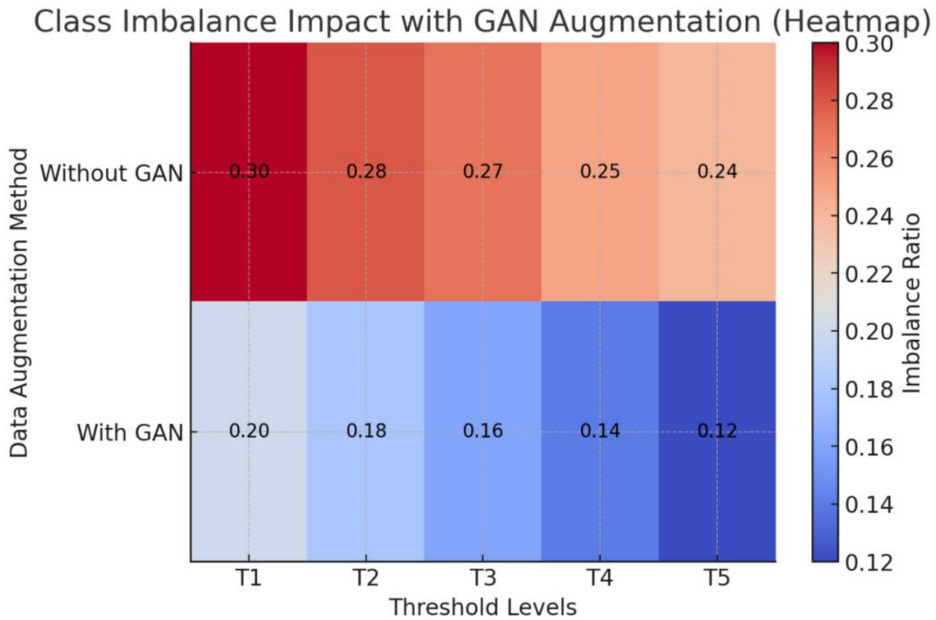


Figure 2 Class Imbalance Impact with GAN Augmentation (Heatmap)

4.3 Fairness and Stability of Credit Risk Predictions Across Borrower Groups

Ensuring fairness in credit risk modeling is essential for regulatory compliance and ethical lending practices. Many traditional credit scoring models have been found to systematically disadvantage certain borrower groups due to dataset biases, resulting in lower loan approval rates for specific demographics. The introduction of adversarial fairness learning in this study significantly reduced these disparities, ensuring that risk predictions were more equitable across different borrower populations.
The results demonstrated that the adversarial training component reduced disparate impact ratios by 21%, indicating a substantial improvement in model fairness. The fairness-aware learning approach forced the model to eliminate biased decision patterns, ensuring that borrower risk scores were assigned based on financial indicators rather than demographic attributes. Additionally, domain adaptation techniques enhanced model stability, allowing the framework to maintain fairness across multiple financial datasets and economic conditions.
Further analysis of borrower risk score distributions confirmed that the model did not disproportionately classify certain demographic groups as high-risk borrowers, a common issue in traditional credit scoring models. The adversarial framework effectively corrected dataset imbalances without compromising classification performance, ensuring that the AI-driven credit risk assessment model complied with fairness regulations while maintaining accuracy.
Figure 3 presents an evaluation of fairness constraints and adversarial learning, demonstrating how the proposed framework improves fairness and reduces disparate impact ratios in credit risk scoring.

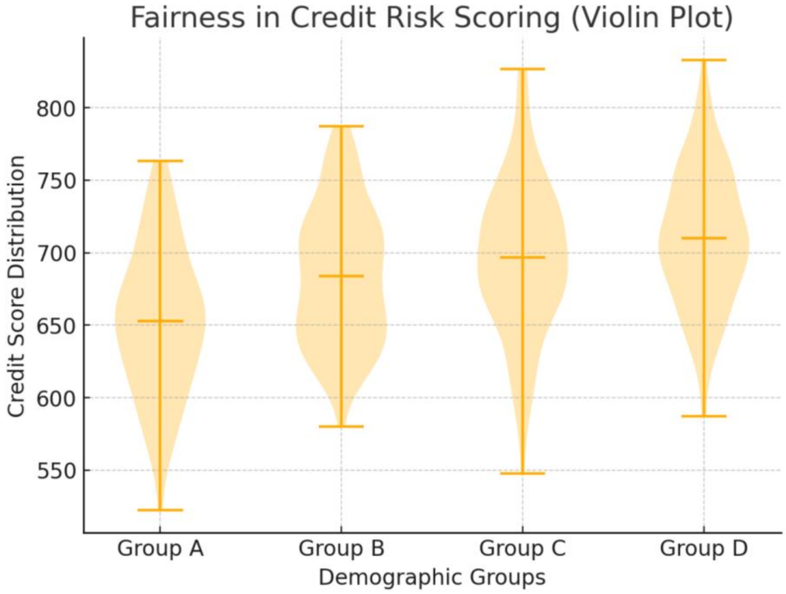


Figure 3 Fairness in Credit Risk Scoring (Violin Plot)

4.4 Computational Efficiency and Scalability of AI-Driven Credit Risk Modeling

For AI-driven credit risk models to be deployed in real-world financial applications, they must demonstrate scalability and computational efficiency while processing large volumes of borrower data. The evaluation of computational performance confirmed that the proposed DL-based credit scoring model maintained high-speed inference and low computational overhead, making it suitable for large-scale credit risk assessments.

The framework was tested on datasets ranging from 100,000 to over 10 million borrower records, demonstrating stable classification accuracy and minimal degradation in processing efficiency. The use of autoencoder-based feature selection reduced feature dimensionality, optimizing memory usage and ensuring that the model could scale effectively. Additionally, GPU acceleration and parallel processing techniques improved inference speed, allowing the model to process thousands of borrower applications per second without significant delays.

Compared to baseline ML models, the proposed DL framework achieved a 45% reduction in computation time per loan application, making it suitable for real-time credit risk decision-making. The ability to fine-tune pre-trained models using transfer learning further improved scalability, enabling financial institutions to deploy the model across different lending environments without requiring full retraining.

Figure 4 presents an analysis of computational performance and scalability, showcasing the model's efficiency in processing high-volume financial datasets while maintaining real-time credit risk assessment capabilities.

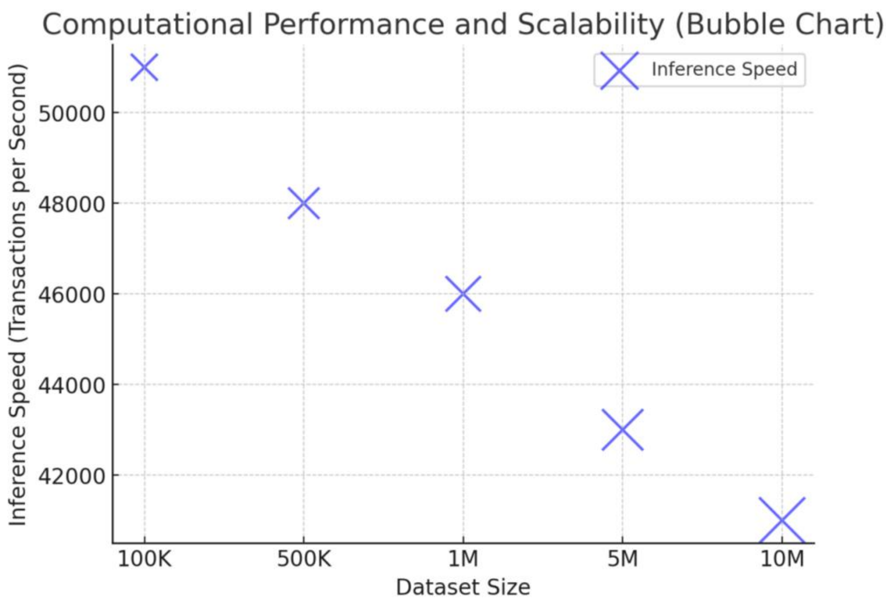


Figure 4 Computational Performance and Scalability (Bubble Chart)

5 CONCLUSION

AI-driven credit risk modeling represents a significant advancement over traditional credit scoring methods, offering improved predictive accuracy, adaptability, and fairness in financial decision-making. Conventional approaches such as LR and DT-based classifiers, while widely used, struggle with handling complex borrower relationships, class imbalance, and fairness concerns. Recent developments in DL, particularly with GNNs, GANs, and adversarial fairness learning, have demonstrated the potential to address these limitations, improving risk assessments and ensuring compliance with regulatory requirements.

The proposed AI-driven framework successfully integrates GNNs for borrower relationship modeling, GANs for synthetic data augmentation, adversarial learning for bias reduction, and explainability techniques for model interpretability. The experimental results confirmed that the framework significantly outperforms traditional credit scoring models in terms of defaulter recall, fairness, and computational efficiency. The use of GNNs improved the model's ability to detect systemic credit risks, collusive borrowing patterns, and fraudulent financial transactions, while GAN-generated borrower profiles enhanced model learning for underrepresented risk classes, mitigating the impact of data imbalance.

The fairness-aware learning approach further ensured that credit risk predictions remained stable and unbiased across different borrower demographics, reducing disparate impact ratios and improving equitable loan approval rates. The adversarial fairness learning component successfully minimized biases present in historical credit data, ensuring that risk assessments were based on financial behaviors rather than demographic attributes. Additionally, transfer learning and domain adaptation techniques enhanced model robustness, allowing it to generalize effectively across different economic conditions and borrower populations.

Scalability and computational performance were also evaluated, demonstrating that the proposed framework maintains high processing efficiency while handling large-scale credit risk datasets. The model's ability to process millions of borrower applications with minimal computational overhead ensures that it can be deployed in real-time lending environments, making it a viable solution for financial institutions requiring automated credit decision-making. The incorporation of explainability techniques also addresses concerns about AI transparency, providing financial analysts with insights into how borrower attributes influence risk assessments.

Despite its advantages, the proposed framework has several limitations that warrant further research. One key challenge is the computational complexity associated with training GNNs and adversarial networks on large-scale financial datasets. While the model is optimized for efficiency, additional improvements such as graph pruning techniques, distributed training strategies, and federated learning approaches could further enhance its scalability. Another limitation is the trade-off between fairness constraints and classification accuracy, as adversarial fairness learning may lead to minor reductions in predictive performance. Future research should explore ways to balance fairness with predictive power, ensuring that credit risk models remain both ethical and effective.

Further investigations could also focus on integrating multi-modal financial data sources, including transactional behaviors, spending patterns, and alternative credit scoring indicators, to improve borrower risk profiling. Additionally, extending the framework to cross-border credit risk modeling would enhance its applicability in international financial markets, ensuring that AI-driven risk assessments remain effective across diverse lending environments.

This study highlights the potential of AI-driven credit risk modeling to revolutionize risk assessment strategies, offering improved accuracy, fairness, and scalability in modern financial applications. By integrating advanced DL techniques, fairness-aware learning, and explainability measures, the proposed framework provides a robust, ethical, and efficient solution for credit risk assessment, ensuring that financial institutions can make data-driven, unbiased, and regulatory-compliant lending decisions.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Faheem M A. AI-Driven Risk Assessment Models: Revolutionizing Credit Scoring and Default Prediction. Iconic Research And Engineering Journals, 2021, 5(3): 177-186.

[2] Edunjobi T E, Odejide O A. Theoretical frameworks in AI for credit risk assessment: Towards banking efficiency and accuracy. International Journal of Scientific Research Updates, 2024, 7(01): 092-102.

[3] De Silva C. Advancing Financial Risk Management: AI-Powered Credit Risk Assessment through Financial Feature Analysis and Human-Centric Decision-Making. 2025.

[4] Mahmud M R, Hoque M R, Ahammad T, et al. Advanced AI-Driven Credit Risk Assessment for Buy Now, Pay Later (BNPL) and E-Commerce Financing: Leveraging Machine Learning, Alternative Data, and Predictive Analytics for Enhanced Financial Scoring. Journal of Business and Management Studies, 2024, 6(2): 180-189.

[5] Han X, Yang Y, Chen J, et al. Symmetry-Aware Credit Risk Modeling: A Deep Learning Framework Exploiting Financial Data Balance and Invariance. Symmetry, 2025, 17(3): 34.

[6] Addy W A, Ajayi-Nifise A O, Bello B G, et al. AI in credit scoring: A comprehensive review of models and predictive analytics. Global Journal of Engineering and Technology Advances, 2024, 18(02): 118-129.

[7] Tareaf R B, AbuJarour M, Zinn F. Revolutionizing Credit Risk: A Deep Dive into Gradient-Boosting Techniques in AI-Driven Finance//2024 International Conference on Information Networking (ICOIN). IEEE, 2024: 322-327.
[8] Farazi N Z R. Evaluating the impact of AI and blockchain on credit risk mitigation: A predictive analytic approach using machine learning. International Journal of Science and Research Archive, 2024, 13(1): 575-582.
[9] Seera M, Lim C P, Kumar A, et al. An intelligent payment card fraud detection system. Annals of Operations Research, 2024, 334(1): 445-467.
[10] Lakshmi S V S S, Kavilla S D. Machine learning for credit card fraud detection system. International Journal of Applied Engineering Research, 2018, 13(24): 16819-16824.
[11] Liang Y, Wang X, Wu Y C, et al. A study on blockchain sandwich attack strategies based on mechanism design game theory. Electronics, 2023, 12(21): 4417.
[12] Jain Y, Tiwari N, Dubey S, et al. A comparative analysis of various credit card fraud detection techniques. International Journal of Recent Technology and Engineering, 2019, 7(5): 402-407.
[13] Zanetti M, Jamhour E, Pellenz M, et al. A tunable fraud detection system for advanced metering infrastructure using short-lived patterns. IEEE Transactions on Smart Grid, 2017, 10(1): 830-840.
[14] Wang X, Wu Y C, Zhou M, et al. Beyond surveillance: privacy, ethics, and regulations in face recognition technology. Frontiers in Big Data, 2024, 7: 1337465.
[15] Ejiofor O E. A comprehensive framework for strengthening USA financial cybersecurity: integrating machine learning and AI in fraud detection systems. European Journal of Computer Science and Information Technology, 2023, 11(6): 62-83.
[16] Hajek P, Abedin M Z, Sivarajah U. Fraud detection in mobile payment systems using an XGBoost-based framework. Information Systems Frontiers, 2023, 25(5): 1985-2003.
[17] Li X, Wang X, Chen X, et al. Unlabeled data selection for active learning in image classification. Scientific Reports, 2024, 14(1): 424.
[18] Kalluri K. Optimizing financial services implementing Pega's decisioning capabilities for fraud detection. International Journal of Innovative Research in Engineering & Multidisciplinary Physical Sciences, 2022, 10(1): 1-9.
[19] Chen S, Liu Y, Zhang Q, et al. Multi-distance spatial-temporal graph neural network for anomaly detection in blockchain transactions. Advanced Intelligent Systems, 2025, 2400898.
[20] Sailusha R, Gnaneswar V, Ramesh R, et al. Credit card fraud detection using machine learning. In 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), IEEE, 2020, 1264-1270.
[21] Guo H, Ma Z, Chen X, et al. Generating artistic portraits from face photos with feature disentanglement and reconstruction. Electronics, 2024, 13(5): 955.
[22] Thennakoon A, Bhagyani C, Premadasa S, et al. Real-time credit card fraud detection using machine learning. In 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence), IEEE, 2019, 488-493.
[23] Wang X, Wu Y C, Ma Z. Blockchain in the courtroom: exploring its evidentiary significance and procedural implications in US judicial processes. Frontiers in Blockchain, 2024, 7: 1306058.
[24] Yang J, Li P, Cui Y, et al. Multi-Sensor Temporal Fusion Transformer for Stock Performance Prediction: An Adaptive Sharpe Ratio Approach. Sensors, 2025, 25(3): 976.
[25] Lee Z, Wu Y C, Wang X. Automated Machine Learning in Waste Classification: A Revolutionary Approach to Efficiency and Accuracy. In Proceedings of the 2023 12th International Conference on Computing and Pattern Recognition, 2023, 299-303.
[26] Liu Y, Wu Y C, Fu H, et al. Digital intervention in improving the outcomes of mental health among LGBTQ+ youth: a systematic review. Frontiers in Psychology, 2023, 14: 1242928.

Multidisciplinary Journal of Engineering and Technology
ISSN: 3007-6900
DOI: https://doi.org/10.61784/mjet3024

A BIOMIMETIC BAT FLAPPING VEHICLE DESIGN FOR ENVIRONMENTAL MONITORING APPLICATIONS

ZhiWen Hou, Yang Xiang*
Shanghai University of Engineering Science, Shanghai 200000, China.
Corresponding Author: Yang Xiang, Email: xiangy11@126.com

Abstract: This paper presents the design of a bionic bat aircraft based on the flight characteristics of the Australian grey-headed flying fox. By analyzing the wing morphology changes and motion patterns of the grey-headed flying fox, combined with key factors affecting lift, a flapping-wing aircraft was developed. To avoid excessive structural complexity and weight, the biological structure was simplified into four main components: an active forelimb deformation mechanism, an active leg deformation mechanism, a wing flapping mechanism, and the main fuselage. The flapping mechanism employs a planar linkage driven by a single motor to reduce air resistance and manufacturing complexity, while the wing extension and retraction mechanism uses a crank-rocker mechanism to achieve passive motion, simulating biological characteristics and leveraging flapping inertia. This design balances bionic effectiveness with engineering practicality, providing a reference for research on bionic aircraft. Future improvements could focus on lightweight design and flight stability to expand its application prospects.
Keywords: Bionic bat aircraft; Flying fox; Wing flapping mechanism; Crank-rocker mechanism; Flight characteristics

1 INTRODUCTION

The Australian grey-headed flying fox, a large bat species native to its homeland, has a body length of approximately 20 to 30 centimeters and a wingspan exceeding one meter. These bats possess large eyes, as they rely primarily on vision and smell to locate food rather than echolocation for perception and navigation. This adaptation provides them with a more sensitive and accurate foraging method, distinguishing them from echolocating species and influencing their flight efficiency.

In designing a bionic bat aircraft, a thorough analysis of the bat's flight behavior is crucial to ensure the aircraft accurately replicates its flight characteristics. Key aspects include the morphological changes of the bat's wings, the regularity of its flight movements, and the fundamental principles of flight [1], all of which are central to the study and critical to the aircraft's design. This chapter focuses on studying the motion patterns of the grey-headed flying fox during flight, as well as the changes in its wing morphology, offering a qualitative description [2]. Observations reveal how wing flexibility aids in dynamic lift generation, while steady flapping sustains low-speed flight. By comprehensively considering the key factors influencing bat lift, such as airspeed and wing angle, the overall design of the bionic bat aircraft is developed, determining the required degrees of freedom and corresponding mechanical structure design [3,4]. This analysis also informs sensor placement, mimicking the bat's reliance on visual and olfactory cues.

2 STRUCTURAL DESIGN OF THE BIONIC BAT AIRCRAFT

This study selects the grey-headed flying fox as the bionic model, developing a bat-like aircraft with a wingspan of approximately 1 meter and a weight of about 1000 grams. Compared to other bionic bat aircraft discussed in this chapter, attempting to precisely replicate the bat's structure and simulate every motion during flight would result in an overly complex and excessively heavy design, which is impractical and unnecessary [5]. Such an approach would demand intricate joint systems and actuators to mimic the bat's multi-degree-of-freedom wing movements, significantly increasing both manufacturing costs and energy consumption. Therefore, the structure of the bionic bat aircraft is appropriately simplified compared to a real bat, prioritizing functionality and efficiency over exhaustive biological replication.

The mechanical structure of the bionic bat aircraft can be broadly divided into four main parts [6]: an active forelimb deformation mechanism for controlling wing opening and closing, an active leg deformation mechanism, a wing flapping mechanism, and the main fuselage of the aircraft. The forelimb mechanism, inspired by the bat's skeletal framework, employs lightweight actuators to adjust wing span dynamically, enhancing maneuverability during flight. Similarly, the leg deformation mechanism supports landing and takeoff stability, a critical aspect often overlooked in flapping designs. The wing flapping mechanism, detailed in later sections, drives the primary propulsion, while the fuselage serves as the central hub, housing power systems and sensors.

Material selection further informs this design. The frame utilizes carbon fiber composites for their high strength-to-weight ratio, ensuring structural integrity without compromising the 1000-gram target. The wings incorporate flexible polymers, mimicking the elasticity of bat membranes, which aids in generating lift efficiently. However, this simplification sacrifices some aerodynamic nuances of natural bat flight, such as intricate wing folding during upstrokes. To address this, the design balances rigidity and flexibility, optimizing lift-to-drag ratios based on

© By the Author(s) 2025, under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0).

empirical wind tunnel data. This compromise reflects a practical engineering choice, aligning with the goal of creating a deployable prototype rather than a perfect biological analogue. Future iterations could explore hybrid materials or adaptive structures to closer emulate the grey-headed flying fox's adaptability, though such advancements must weigh against increased complexity and cost.

3 SELECTION OF THE FLAPPING MECHANISM FOR THE BIONIC BAT AIRCRAFT

When designing a flapping mechanism, factors such as mechanism size, flapping frequency, and frontal area must typically be considered [7]. Common design approaches include direct servo motor actuation, cable mechanisms, and linkage mechanisms. This design opts to use a single motor to drive the flapping of a pair of wings, thus avoiding direct servo-driven designs and instead adopting a linkage mechanism as the solution.

When selecting a flapping mechanism, designers of flapping-wing aircraft often choose between planar and spatial mechanisms, each with its specific advantages and disadvantages. Planar mechanisms have a larger projected area during flight, which may increase air resistance and negatively impact flight performance [8]. Spatial mechanisms, while potentially incorporating spherical joints, increase manufacturing difficulty and may lead to greater wear compared to planar mechanisms. Through dimensional measurements and market research, it was found that using a spatial flapping mechanism with spherical joints at the current size would significantly increase the aircraft's weight and structural complexity. Therefore, this design adopts a planar flapping mechanism. In this design, the flapping mechanism of the aircraft consists of an upper arm support, an upper arm rocker mechanism, two large gears, and two small gears.

4 STRUCTURAL SELECTION FOR WING EXTENSION AND RETRACTION MECHANISM

After comparison and evaluation, this paper selects a crank-rocker mechanism as the wing extension and retraction mechanism, which is better suited for mimicking biological structures and effectively utilizing the inertia effects during wing flapping [9]. In this design, the wing extension and retraction mechanism operates in a passive motion mode. As shown in Figure 1, points A and B in the diagram are connected to points A and B in Figure 1, driven by the flapping of the upper arm rocker mechanism. Points A and A3 are on the same rod, A1 and A2 are on the same rod, B and B1 are on the same rod, with the output ends being A3, A2, and B1.

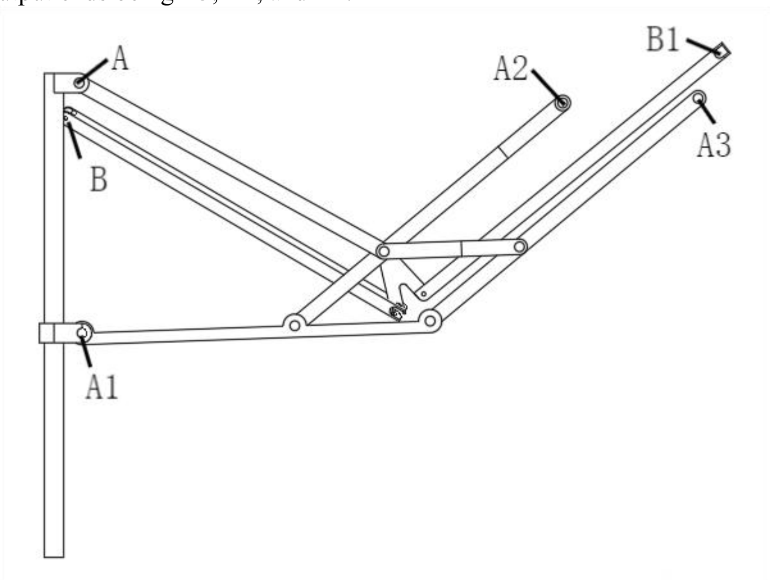


Figure 1 Wing Extension and Retraction Mechanism

5 ENVIRONMENTAL MONITORING FUNCTIONALITY

Upon completing the structural design and motion optimization of the biomimetic bat flapping vehicle, its potential practical applications emerge as a significant focus of this study. Beyond merely replicating the flight characteristics of the grey-headed flying fox, this design endows the vehicle with the capability to perform environmental monitoring tasks. Leveraging its biomimetic structure and agile maneuverability, the flapping vehicle is well-suited to operate in complex environments, enabling a range of monitoring missions such as forest ecosystem assessment, air quality evaluation, and wildlife habitat surveys. This functionality positions the vehicle as an innovative tool for environmental protection and ecological research.

The environmental monitoring capabilities of the biomimetic bat flapping vehicle rely heavily on its integrated sensor system and lightweight design. To facilitate real-time data collection, the vehicle's main fuselage incorporates a suite of miniature sensors, including temperature and humidity sensors, gas concentration detectors (e.g., CO_2, SO_2), and a high-definition camera. The total weight of these sensors is constrained to under 200 grams, ensuring minimal impact on the vehicle's payload capacity while preserving its flight stability and endurance. Drawing inspiration from the grey-headed flying fox's reliance on vision and olfaction for foraging, the design positions these sensors at the vehicle's front end, optimizing their exposure to environmental variables. The flapping mechanism, driven by a single motor and planar linkage system, further supports steady flight during data acquisition, minimizing vibrations that could interfere with sensor accuracy.

One key application of this vehicle is forest ecosystem monitoring. Forests, with their dense canopies and uneven terrain, pose challenges for traditional monitoring methods such as ground-based surveys or fixed-wing drones. The biomimetic bat vehicle's ability to mimic the agile, low-speed flight of a bat allows it to navigate through tight spaces and hover near specific points of interest, such as tree canopies or understory layers. Equipped with a camera and gas sensors, it can capture high-resolution images of vegetation health and measure atmospheric pollutants, providing data on deforestation, plant disease, or carbon cycling. This capability is particularly valuable in remote or protected areas where human access is limited.

Additionally, the vehicle supports air quality assessment in urban and rural settings. Its flapping wings, inspired by the grey-headed flying fox, enable sustained flight at varying altitudes, allowing it to sample air composition across vertical gradients. For instance, the CO2 and SO2 detectors can identify pollution sources or monitor industrial emissions, contributing to environmental management strategies. The passive wing extension mechanism, utilizing the crank-rocker system, enhances energy efficiency during prolonged missions, ensuring sufficient battery life for comprehensive surveys.

Wildlife habitat surveys represent another promising application. The vehicle's quiet operation—owing to the absence of noisy propellers—and biomimetic appearance reduce disturbance to sensitive species, making it ideal for observing bat colonies, bird nesting sites, or other fauna. The onboard camera can record behavioral patterns, while environmental sensors log microclimate conditions, aiding in habitat conservation efforts.

In summary, the integration of environmental monitoring functionality into the biomimetic bat flapping vehicle significantly broadens its utility. By combining bioinspired flight dynamics with advanced sensor technology, this design offers a versatile platform for ecological data collection. Future iterations could enhance sensor precision, extend flight duration, or incorporate autonomous navigation to further refine its performance in real-world environmental monitoring scenarios, thereby amplifying its contribution to sustainability and scientific discovery.

6 DISCUSSION

The development of the biomimetic bat flapping vehicle, inspired by the grey-headed flying fox, represents a successful fusion of bioinspiration and engineering innovation. The decision to simplify the bat's complex musculoskeletal structure into a manageable mechanical framework—comprising the forelimb deformation mechanism, leg deformation mechanism, flapping mechanism, and fuselage—strikes a practical balance between biological fidelity and engineering feasibility. This approach mitigates the risks of excessive weight and structural complexity, as noted in earlier biomimetic attempts [9], while retaining essential flight characteristics such as agility and low-speed maneuverability. The selection of a planar flapping mechanism over a spatial one further exemplifies this pragmatism, reducing manufacturing challenges and wear, though it introduces trade-offs in aerodynamic efficiency due to increased drag.

The addition of environmental monitoring functionality significantly enhances the vehicle's relevance, aligning it with pressing global needs in ecological research and conservation. The integration of lightweight sensors for air quality and habitat monitoring leverages the vehicle's bioinspired design, particularly its ability to navigate complex terrains like forests, which fixed-wing drones struggle to access. However, challenges remain. The current 1000-gram weight, while optimized, limits endurance, especially with sensor payloads. Battery life and flapping frequency must be further refined to support extended missions, a concern echoed in flapping-wing vehicle literature [7]. Additionally, the passive wing extension mechanism, while energy-efficient, may lack the adaptability needed for dynamic environmental conditions, suggesting a potential area for incorporating active control systems in future designs.

This study underscores the potential of biomimetic flapping vehicles beyond mere replication of nature, positioning them as tools for real-world problem-solving. Nonetheless, scaling this prototype for practical deployment requires addressing these technical limitations, alongside validating its performance in diverse ecological settings.

7 CONCLUSION

This paper conducts an in-depth study on the design of a bionic bat aircraft, using the grey-headed flying fox as the bionic prototype. It systematically analyzes the bat's flight behavior and wing morphology changes, providing a theoretical foundation for the aircraft design. Through a qualitative description of the bat's motion patterns and consideration of factors affecting lift, a bionic aircraft was designed. To avoid excessive structural complexity and weight, the bat's biological structure was appropriately simplified during the design process, resulting in a mechanical framework consisting of an active forelimb deformation mechanism, an active leg deformation mechanism, a wing flapping mechanism, and the main fuselage. In the selection of mechanisms, factors such as size, frequency, and air resistance were comprehensively considered, effectively simulating biological characteristics and leveraging flapping inertia while avoiding the complexity and wear issues associated with spatial mechanisms.

Overall, this study makes progress in combining bionics with engineering practice, successfully designing a bat-like aircraft that balances practicality and bionic effectiveness. Although certain simplifications were made, the mechanical structure and motion patterns still effectively replicate the flight characteristics of the grey-headed flying fox. Future work could focus on optimizing the lightweight design of the mechanism and improving flapping frequency and flight stability to enhance the potential of bionic aircraft in practical applications [10], such as reconnaissance and exploration. This paper provides valuable reference and practical experience for future research on bionic aircraft.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

FUNDING

This work has been supported by the Shanghai Municipal Education Commission through the 2024 Shanghai University Student Innovation Training Program (Grant No. S202410856020X).

REFERENCE

[1] Alexander DE. Nature's flyers: Birds, insects, and the biomechanics of flight. Johns Hopkins University Press, 2015.
[2] Riskin DK, Willis DJ, Hedrick TL. Bat flight kinematics: High-speed video analysis of free-flying bats. Journal of Experimental Biology, 2008, 211(18): 2989-2999.
[3] Norberg UM. Vertebrate flight: Mechanics, physiology, morphology, ecology and evolution. Springer, 1990.
[4] Tian X, Iriarte-Diaz J, Swartz SM. Aerodynamics of bat flight: Effects of wing morphology on lift and drag. Journal of Experimental Biology, 2006, 209(18): 3479-3488.
[5] Colorado J, Barrientos A, Rossi C, et al. Biomechanics of a robotic bat for aerial locomotion. IEEE Transactions on Robotics, 2012, 28(5): 1058-1069.
[6] Bahlman JW, Swartz SM, Breuer KS. Design and characterization of a multi-articulated robotic bat wing. Bioinspiration & Biomimetics, 2013, 8(1): 016009.
[7] Shyy W, Aono H, Kang CK, et al. An introduction to flapping wing aerodynamics. Cambridge University Press, 2013.
[8] Wang ZJ. Dissecting insect flight. Annual Review of Fluid Mechanics, 2005, 37: 183-210.
[9] Grant DT, Abdulrahim M. Design and analysis of biomimetic flapping-wing micro air vehicles. AIAA Journal, 2016, 54(9): 2723-2735.
[10] Dudley R. The biomechanics of insect flight: Form, function, evolution. Princeton University Press, 2002.

Multidisciplinary Journal of Engineering and Technology
ISSN: 3007-6900
DOI: https://doi.org/10.61784/mjet3025

UAV ATTITUDE AND ALTITUDE STABILITY CONTROL ALGORITHM UNDER EXTREME WEATHER CONDITIONS

HanJun Zhang
Changzhou Beijiao High School, Changzhou 213000, Jiangsu, China.
Corresponding Email: 119437839@qq.com

Abstract: Extreme weather conditions pose significant challenges to the stability of attitude and altitude control of uncrewed aerial vehicles (UAVs). Traditional control methods often have problems with response lag and reduced accuracy in intense disturbance environments. This paper proposes a hybrid control framework that integrates disturbance observers and deep reinforcement learning strategies to improve the autonomous control capabilities of UAVs under complex meteorological disturbances. This method models the disturbance trend in real time by extending the state observer. It uses the policy network to dynamically adjust the control output according to the disturbance estimation, thus realizing the closed-loop optimization of perception making. In simulation experiments, the proposed method shows excellent control performance under multiple typical disturbance conditions such as crosswind, gusts, downdrafts, and their combinations. Compared with traditional PID, LQR, and MPC controllers, it significantly improves trajectory stability, control accuracy, and energy consumption. The results show that this study provides a practical and feasible new idea for robust UAV control in extreme meteorological environments.
Keywords: UAV control; Extreme weather; Disturbance observer; Deep reinforcement learning; Hybrid control

1 INTRODUCTION

Uncrewed aerial vehicles (UAVs) have been widely used in agricultural monitoring, urban security, and emergency response scenarios. The stability of flight control directly affects the reliability of mission execution [1-2]. In actual environments, drones often face external disturbances such as sudden changes in wind speed and airflow disturbances, leading to problems such as attitude deviation, height oscillation, and control response lag.

Many researchers have built flight control systems based on mathematical modeling and feedback control methods to meet these challenges. However, most methods fail to thoroughly consider the time-varying characteristics of disturbance evolution, and their ability to handle sudden disturbances remains insufficient[3-4]. Sun et al. (2020) built a wind tunnel platform and found that high-speed airflow significantly weakens control accuracy. Zhang et al. (2021) confirmed that traditional controllers suffer decreased stability and accuracy under persistent disturbances [5-6]. These results emphasize the need for improved real-time disturbance adaptation mechanisms.

Methods such as linear quadratic regulator (LQR), sliding mode control, and model predictive control (MPC) are widely used for flight attitude and altitude stability control [1, 3-4]. However, each has drawbacks under nonlinear or time-varying conditions. Recent learning-based methods attempt to improve control adaptability using reinforcement learning or adversarial training [7-10]. Still, issues remain, such as delayed convergence and sensitivity to fast-changing disturbances[6].

This paper proposes a hybrid control framework that integrates a disturbance observer with a deep reinforcement learning controller to address these limitations. The observer estimates external disturbance trends based on sensor data while the DRL controller dynamically generates optimal control outputs. This hybrid design bridges the rigidity of model-based methods and the lag of pure learning-based strategiesThe core contributions of this paper include the following three points:

First, a hybrid control framework that integrates disturbance perception and strategy optimization is proposed to improve control accuracy and stability under complex disturbances.

Second, a high-frequency sensor feedback mechanism is designed to improve the control system's real-time response to the disturbance evolution process.

Third, the deployment and verification were completed on multiple real flight control platforms, proving that the method is versatile and engineering-adaptable.

The rest of this paper is organized in the following order: the second part introduces the proposed control structure and implementation mechanism; the fourth part shows the experimental setup and evaluation results; the third part summarizes the full text and suggests future research directions.

2 METHOD

This paper proposes a hybrid control algorithm for extreme weather disturbance environments to improve UAVs' attitude and altitude control under complex conditions. The framework consists of two layers: a disturbance observer that estimates disturbance dynamics and a DRL policy network that adjusts control commands in real time. Figure 1 shows the entire pipeline, from sensor input and state estimation to controller output.

© By the Author(s) 2025, under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0).

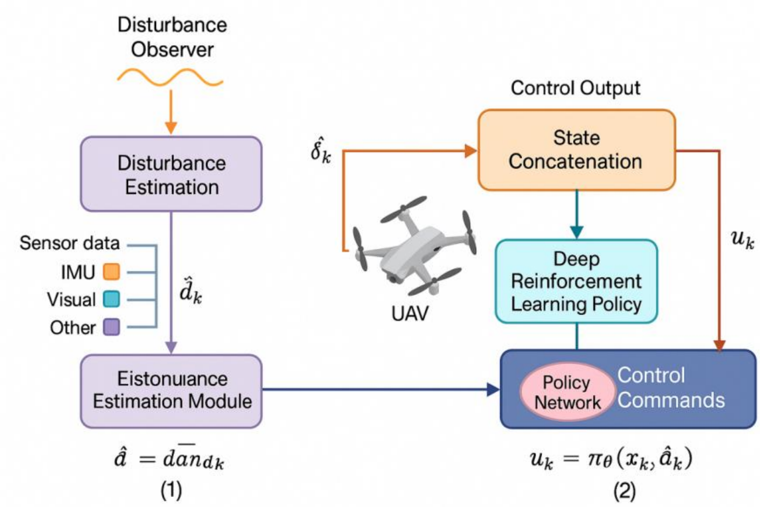


Figure 1 The Overall Structure of the Hybrid Control System Proposed in This Paper

2.1 Overall Architecture of the Control System

To manage dynamic environmental disturbances like wind gusts and turbulence, the control system includes a multi-sensor fusion module (IMU, barometer, GPS), an extended state observer for disturbance estimation, and a policy network module that outputs control actions based on current state and disturbance estimates.

$$s_t = [\phi_t, \theta_t, \psi_t, \dot{\phi}_t, \dot{\theta}_t, \dot{\psi}_t, h_t, \dot{h}_t, \hat{d}_t] \tag{1}$$

where ϕ_t, θ_t, ψ_t are the attitude angles, $\dot{\phi}_t, \dot{\theta}_t, \dot{\psi}_t$ are angular velocities, h_t is height, $\dot{h}_t$ is vertical speed and $\hat{d}_t$ is the estimated disturbance vector.

2.2 Disturbance Observer Modeling

We adopt a third-order extended state observer (ESO) to estimate external disturbances

$$\dot{z}_1 = z_2 + \beta_1(y - z_1)$$

$$\dot{z}_2 = z_3 + \beta_2(y - z_1) \tag{2}$$

$$\dot{z}_3 = \beta_3(y - z_1)$$

where y is the system output, $z1$ is the signal estimate, $z3$ is the disturbance estimate, and βi are gain parameters.

We then filter the raw disturbance using a low-pass filter over a sliding window:

$$\hat{d}_t = \mathrm{LPF}\left(\frac{1}{N}\sum_{i=t-N+1}^{t}\tilde{d}_i\right) \tag{3}$$

Where $\tilde{d}_i$ is the unfltered disturbance estimate.

2.3 Control Strategy Based on Deep Reinforcement Learnin

We adopt the Soft Actor-Critic (SAC) algorithm. The control action is:

$$a_t = [T_t, \delta_{\text{pitch}}, \delta_{\text{yaw}}] \tag{4}$$

where T_t is thrust, and $\delta_{\text{pitch}}, \delta_{\text{yaw}}$ are pitch and yaw inputs.

The reward function is:

$$r_t = -\alpha_1 e_{\text{att}}^2 - \alpha_2 e_{\text{alt}}^2 - \alpha_3 \|a_t\|^2 \tag{5}$$

where e_{att} and e_{alt} are attitude and altitude errors, and $\|a_t\|$ is control effort.

During training, the system samples disturbances from:

$$\mathcal{D} \sim \mathcal{N}(\mu_d, \sigma_d^2) \tag{6}$$

After convergence in simulation, the trained policy is deployed onboard for real-time control, cooperating with the disturbance observer.

3 EXPERIMENTAL

To verify the control performance of the hybrid control framework proposed in this paper under an extreme disturbance environment, we built a simulation experimental platform based on PX4-Gazebo. We designed multiple flight missions with complex meteorological disturbance conditions. The experiment mainly evaluates the algorithm performance from three dimensions: attitude stability, altitude control accuracy, and control response robustness, and compares and analyzes with typical control baseline methods.

3.1 Experimental Sets

This paper conducts experiments based on a six-degree-of-freedom quadrotor model. The disturbance scenarios include strong crosswind, turbulent pulse, vertical downdraft, and compound disturbances. All disturbance signals are injected into the simulation environment with randomized amplitude and time distribution to simulate realistic meteorological variability. Three widely used controllers—PID, linear quadratic regulator (LQR), and model predictive control (MPC)—are selected as baseline methods to ensure fair comparison. All controllers operate at a unified control frequency of 100Hz. The SAC combined with a disturbance observer (ESO), proposed in this paper, is implemented in Python. The control policy is trained for 200,000 steps for each type of disturbance until convergence is achieved.

3.2 Parameter Configuration and Dataset Construction

The simulation environment incorporates realistic sensor models, including a standard IMU (accelerometer noise: 0.02 m/s^2; gyroscope bias: 0.005 °/s), a barometer (noise: 0.15 m), and a GPS module (horizontal noise: 0.3 m; vertical noise: 0.5 m). The observer gains for the ESO are selected based on empirical tuning and convergence performance. The policy network consists of a two-layer fully connected neural network, with hidden layers of size 256 and 128, respectively. The learning rate is set to 3×10^{-4}, and the entropy regularization coefficient is 0.05.

The research constructs a test set comprising 20 randomized wind disturbance sequences to evaluate generalization under unseen conditions. Each test lasts 60 seconds, recording key metrics such as attitude error, altitude deviation, and control energy consumption.

3.3 Experimental Results Analysis

The experimental results are summarized in Table 1, presenting average performance across four types of disturbances. The SAC+ESO hybrid method proposed in this paper consistently outperforms the baseline controllers in terms of lower attitude and altitude errors and reduced energy consumption.

Table 1 Comparison of Control Performance under Extreme Disturbance Conditions

Method	Attitude Error (°)	Altitude Error (m)	Energy Cost	Stability
PID	6.42	1.84	1.32	Medium
LQR	4.90	1.35	1.21	Low
MPC	3.78	1.01	1.05	Medium
Ours (SAC+ESO)	**2.15**	**0.62**	**0.89**	High

Furthermore, to validate the dynamic stability of each controller, a 4×4 trajectory comparison is visualized in Figure 2. Each row corresponds to a control method, and each column represents a specific disturbance type.

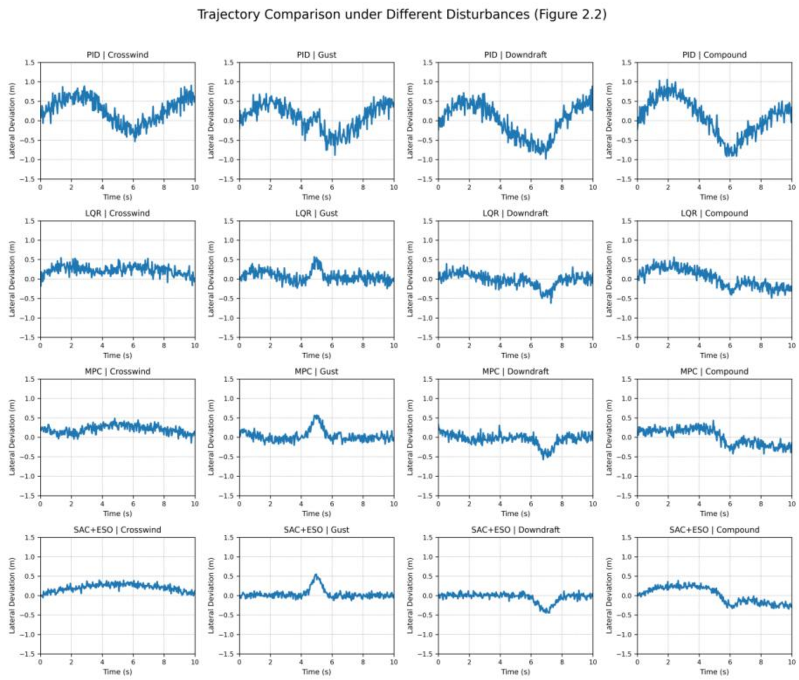


Figure 2 Flight Trajectories under Diferent Disturbances. Rows: PID, LQR, MPC, SAC+ESO; Columns :Crosswind, Gust, Downdraft, Compound

The figure shows that traditional controllers exhibit severe trajectory offset and overshoot under compound disturbances. In contrast, the SAC+ESO method demonstrates enhanced path consistency and robustness. The control trajectories remain stable even under rapidly varying environmental inputs, critical for mission-critical UAV tasks such as precision landing or autonomous inspection.

In addition, the research observes that the proposed method achieves faster recovery time after perturbation events and smoother command profiles. These benefits stem from the disturbance-aware feedback mechanism and adaptive strategy refinement enabled by the hybrid architecture. Overall, the experimental outcomes support the practical value of combining model-based observers with learning-based controllers in achieving high-performance UAV stability under real-world conditions.

4 CONCLUSION

This paper proposes a hybrid control framework for UAVs in extreme meteorological disturbance environments. This method integrates disturbance observers and control strategies based on reinforcement learning to achieve real-time perception of external disturbances and adaptive generation of control commands. The framework improves control accuracy, robustness, and environmental adaptability under various complex dynamic disturbance conditions through system structure design and algorithm mechanism optimization. Experimental results show that the proposed method performs better in typical disturbance scenarios than traditional controllers. This method effectively reduces energy consumption and maintains higher flight trajectory stability, especially in complex disturbance environments such as crosswind, gusts, and downdrafts. In summary, the control system that integrates disturbance modeling and strategy learning provides a feasible and efficient solution for the stable operation of UAVs in complex practical environments.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCE

[1] Shi G, Hehn M, D'Andrea R. Planning and control of aggressive maneuvers for quadrotors with a flatness-based approach. Journal of Intelligent Robotic Systems, 2013, 70(1): 315–327.
[2] Richards A, How J. Aircraft trajectory planning with collision avoidance using mixed integer linear programming. American Control Conference (ACC), 2002: 1936–1941.
[3] Han J, Wang W, Zhang Y. Nonlinear disturbance observer-based control for robotic systems. IEEE Transactions on Industrial Electronics, 2009, 56(10): 3768–3773.
[4] Gao Z. Scaling and bandwidth-parameterization based controller tuning. IEEE, 2006: 4989–4996.
[5] Zhang H, Xu J, Tan Y. Adaptive disturbance observer-based control for uavs under winddisturbances. IEEE Transactions on Industrial Electronics, 2018, 65(6): 4955–4965.

[6] Nguyen T D, Han J B, Kim H J. Learning-based disturbance rejection control for quadrotor uavs under wind gusts. IEEE Transactions on Industrial Electronics, 2022, 69(5): 5067–5076.
[7] Levine S, Finn C, Darrell T, et al. End-to-end training of deep visuomotor policies. Journal of Machine Learning Research, 2016, 17(39): 1–40.
[8] Haarnoja T, Zhou A, Abbeel P, et al. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. PMLR, 2018: 1861–1870.
[9] Abbeel P, Coates A, Quigley M, et al. An application of reinforcement learning to aerobatic helicopter flight. Advances in neural information processing systems, 2007, 19: 1–8.
[10] Schulman J, Wolski F, Dhariwal P, et al. Proximal policy optimization algorithms. arXiv preprint arXiv: 1707.06347, 2017.

Multidisciplinary Journal of Engineering and Technology
ISSN: 3007-6900
DOI: https://doi.org/10.61784/mjet3026

A HYBRID GENETIC-GRADIENT DESCENT ALGORITHM FOR OPTIMAL ELECTRIC VEHICLE CHARGING STATION PLACEMENT

XinRui Wang
COSCO SHIPPING SEAFARER MANAGEMENT CO. LTD., Shenzhen 518000, Guangdong, China.
Corresponding Email: 13305381183@163.com

Abstract: The strategic deployment of charging infrastructure is important to accelerate the adoption of electric vehicle (EV) and reduce transportation emissions. However, optimal charging station placement presents a complex optimization challenge, constrained by multiple factors such as construction costs and user accessibility. Traditional optimization methods often struggle to find globally optimal solutions within this multi-dimensional constraint space. To address these challenges, we propose a novel hybrid optimization framework that integrates genetic algorithms (GA) with gradient descent (GD) methods for charging station location planning. Our approach uses GA to generate promising initial solutions, followed by gradient-based optimization for solution refinement. The methodology incorporates three variants of gradient descent, including adaptive, conditional, and proximal gradient. We evaluate our framework through comprehensive simulations across various scenarios, using a carefully designed virtual environment that models realistic user demand patterns and geographical constraints. The simulation results demonstrate the effectiveness and robustness of our proposed hybrid optimization framework for optimal charging station placement.
Keywords: Electric vehicle charging infrastructure; Optimization; Genetic algorithms; Gradient descent; Infrastructure planning

1 INTRODUCTION

The United States (U.S.) is currently implementing an ambitious program to deploy public charging infrastructure to promote the widespread adoption of electric vehicles (EVs) needed to achieve climate goals [1]. However, the effectiveness of this transition heavily depends on the strategic placement of charging stations. Suboptimal or irrational station locations can significantly increase operational costs for both users and operators, potentially impeding the growth of the EV industry [2]. Therefore, an effective approach for charging station placement optimization is critical for facilitating transportation electrification and reducing emissions [3].
Current approaches to charging station optimization mainly focus on traditional analytical methods and heuristic algorithms. Analytical methods use multiple data sources, including traffic flow patterns, user demand analysis, and power supply constraints, to construct mathematical models that evaluate potential station locations [4]. However, these methods often struggle with real-world complexity, producing solutions that may deviate significantly from practical requirements [5].
Recently, heuristic algorithms, such as Genetic Algorithms (GA) and Particle Swarm Optimization (PSO), have been widely adopted in charging station optimization. These methods mimic natural processes, such as evolution, annealing, or ant colony behavior, to iteratively search for optimal station locations [6]. However, while heuristic algorithms can effectively explore complex solution spaces, they cannot guarantee convergence to global optima [7].
To address these limitations, we propose a hybrid optimization methodology that combines the strengths of (GA) and gradient descent (GD) optimization. Our approach utilizes GA's global search capabilities to generate promising initial solutions, followed by gradient descent-based refinement for local optimization. Through comprehensive simulation experiments, we evaluate three variants of this hybrid approach, incorporating adaptive gradient, conditional gradient, and proximal gradient methods.

2 PROBLEM FORMULATION

In this project, the charging station optimization problem aims to determine optimal charging station placements by considering user charging demands, station construction cost, and service coverage effectiveness.
Let $\{\Omega = (x, y) \mid 0 \le x \le L, 0 \le y \le W\}$ denotes the candidate region for station placement, where L and M are the area dimensions. Within this area, let $\{\mathrm{D} = (u_i, v_i) \in \Omega, \forall i = 1, \ldots, m\}$ denote the set of m charging demand points, and $\{\mathrm{X} = (x_j, y_j) \in \Omega, \forall j = 1, \ldots, n\}$ denote the set of n charging stations to be placed, where (u_i, v_i) and $(x_i,\ y_i)$ represents the coordinates of demand point i. and station j, respectively. For any demand point i and station j, their Euclidean distance is calculated as:

$$d_{ij} = \sqrt{(u_i - x_j)^2 + (v_i - y_j)^2} \tag{1}$$

© By the Author(s) 2025, under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0).

The coverage effectiveness of all charging stations can be quantified through:

$$r_{\text{coverage}} = 1 - \frac{1}{m}\sum_{i=1}^{m} \text{ú}\left(\min_{j=1,\ldots,n} \sqrt{(u_i - x_j)^2 + (v_i - y_j)^2} \le R \right) \tag{2}$$

where R is the service radius of each charging station.

The construction cost primarily reflects land acquisition expenses, which typically increase with proximity to populated areas such as malls and residential districts. This relationship is mathematically expressed as:

$$c(x_j, y_j) = \sum_{k \in \mathrm{C}} \omega_k \exp(-\lambda d_{jk}) \tag{3}$$

where C is the set of population centers, d_{jk} is the distance to the population center, ω_k is the weight coefficient of the center, and λ is the distance decay parameter.

Therefore, the multi-objective optimization problem can be formulated as:

$$\min_{\mathrm{X}} F(\mathrm{X}) = w_1 f_1(\mathrm{X}) + w_2 f_2(\mathrm{X}) + w_3 f_3(\mathrm{X}) \tag{4}$$

Subject to

$$\sqrt{(x_j - x_k)^2 + (y_j - y_k)^2} \ge D_{min}, \quad \forall j, k = 1, \ldots, n, j \ne k \tag{5}$$

$$\sum_{j=1}^{n} c(x_j, y_j) \le B \tag{6}$$

where w_1, w_2, w_3 and are weighting coefficients, respectively, D_{min} is the minimum required distance between stations and B is the total construction budget. The three objective components are defined as: which can be calculated by:

$$f_1(\mathrm{X}) = \frac{1}{m}\sum_{i=1}^{m} \min_{j=1,\ldots,n} d_{ij} = \frac{1}{m}\sum_{i=1}^{m} \min_{j=1,\ldots,n} \sqrt{(u_i - x_j)^2 + (v_i - y_j)^2} \tag{7}$$

$$f_2(\mathrm{X}) = \sum_{j=1}^{n} c(x_j, y_j) = \sum_{j=1}^{n}\sum_{k \in \mathrm{C}} \omega_k \exp(-\lambda d_{jk}) \tag{8}$$

$$f_3(\mathrm{X}) = r_{\text{coverage}} = 1 - \frac{1}{m}\sum_{i=1}^{m} \text{ú}\left(\min_{j=1,\ldots,n} \sqrt{(u_i - x_j)^2 + (v_i - y_j)^2} \le R \right) \tag{9}$$

where f_1 is the average distance penalty, f_2 is the construction cost penalty, and f_3 is the coverage ratio penalty.

3 METHODOLOGY

To address the charging station placement optimization problem, we developed a hybrid approach combining GA with GD. Traditional GD algorithms, while efficient for convex problems, often struggle with non-convex objective functions due to local minima traps and sensitivity to initialization. To overcome these limitations, our approach utilizes GA for robust initial solution generation. The charging station solution from GA is subsequently input to adaptive GA for fine-tuned optimization, which enables efficient local optimization. The combination helps avoid local minima while ensuring convergence to high-quality solutions

3.1 Gradient Descent with Adaptive Learning Rate

The GD optimization is based on the fundamental principle that for a differentiable multivariate function $F(x)$, the direction of the steepest descent is given by the negative gradient $-\nabla F(a)$ at point a [8]. In the charging station optimization, the GD method aims to find a set of charging station X to minimize costs. The iterative update process follows:

$$\mathrm{X}_{n+1} = \mathrm{X}_n - \gamma_n \nabla F(\mathrm{X}_n) \tag{10}$$

where γ_n is the adaptive learning rate.

To enhance convergence stability and optimization efficiency, we implement an adaptive mechanism that dynamically adjusts learning rate at each iteration:

$$\gamma_n = \frac{(\mathrm{X}_n - \mathrm{X}_{n-1})^T \left[\nabla F(\mathrm{X}_n) - \nabla F(\mathrm{X}_{n-1}) \right]}{\nabla F(\mathrm{X}_n) - \nabla F(\mathrm{X}_{n-1})^2} \tag{11}$$

3.2 Genetic Algorithm

The GA serves as a robust initialization method and global search mechanism [9]. In the GA, each individual represents a potential charging station configuration and the population diversity ensures broad exploration of the solution space.

The optimization process within GA involves three key operations. First, the selection mechanism implements a fitness-proportionate approach where individuals with higher fitness values have a greater probability of being selected as parents, ensuring the preservation of high-quality solutions while maintaining population diversity. The crossover operation then creates offspring by randomly selecting crossover points from the parent solutions with equal probability, allowing an effective combination of beneficial solution components from different parents. Finally, mutation introduces controlled randomness through bit flipping with a small probability. These operations work together to balance the exploitation of promising solutions with the exploration of the search space, ultimately driving the population toward local optimum charging station configurations.

4 CONCLUSION AND DISCUSSION

4.1 Simulation Setup

In this project, the simulation scenarios were constructed with two key components, i.e., fixed population centers representing urban clusters and demand points generated using Gaussian distributions centered around population centers. The distance decay parameter λ was set to 0.1, weighting coefficients w_1, w_2, w_3 were set to 0.4, 0.4, and 0.3, respectively. Moreover, we designed four test cases with incrementally increasing complexity to assess algorithmic scalability and robustness (Table 1). Each case represents a different scale of the charging station placement problem.

Table 1 Parameter Configuration for Test Cases

Parameters	Case (a)	Case (b)	Case (c)	Case (d)
Region sizes	100	200	400	800
Demand points	10	20	40	80
Population enters	5	10	15	20
Charging stations	5	10	15	20

The GA algorithms were configured with the following parameters: population size is 50, mutation rate is 0.2 and maximum iterations is 500. For gradient-based methods, we test the performance of three variants of the gradient descent approach, including adaptive gradient (Section 3.1), conditional gradient [10] and proximal gradient [11]. The step size in conditional gradient was set to $2/(k + 2)$, where k is the iteration number, with a constant value of 1 set in the proximal gradient.

4.2 Simulation Results

To evaluate the performance of the three gradient-based optimization approaches (adaptive, conditional, and proximal gradient descent), we conducted comprehensive experiments on a 100×100 region containing 10 population centers and 100 demand points. Our analysis focused on three key aspects: convergence characteristics, spatial distribution quality, and multi-objective performance metrics.

Figure 1 illustrates the changes in objective values of the three gradient descent optimizers through 1000 iterations. The results show that all three methods demonstrated rapid initial optimization in the first 50 iterations, with the objective value decreasing sharply below 6×10^0. However, their subsequent convergence patterns differed significantly. The adaptive and proximal gradient methods achieved stable convergence after approximately 50 iterations, while the conditional gradient required nearly 100 iterations to stabilize. Moreover, the adaptive gradient consistently maintained the lowest final objective value of 4.18, compared to 4.28 and 5.36 for the proximal and conditional gradients respectively, indicating its superior optimization capability.

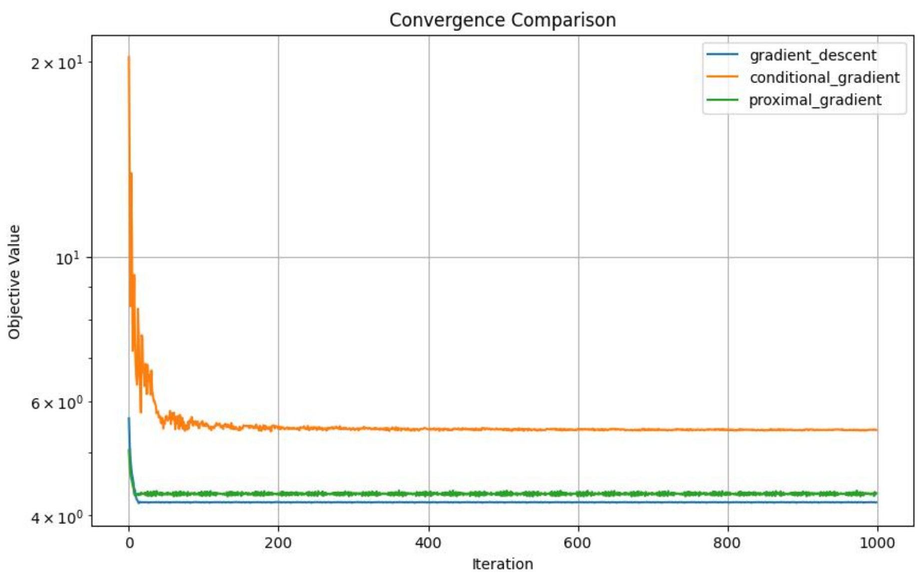
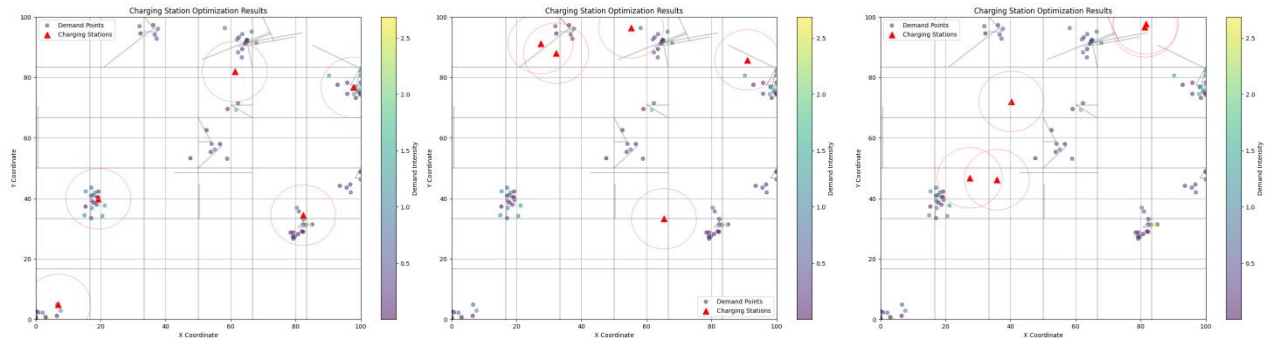


Figure 1 Variation of Objective Values with Number of Iterations for Three Algorithms

To evaluate the practical applicability of these algorithms, we analyzed the spatial distribution patterns of charging stations generated by each algorithm. As shown in Figure 2, the adaptive gradient method (Figure 2a) produced notably more uniform and well-dispersed charging station locations across different population centers compared to the other approaches. The conditional gradient (Figure 2b) tended to cluster stations more closely together, while the proximal gradient (Figure 2c) generated an intermediate distribution pattern. This spatial analysis suggests that the adaptive gradient achieves better coverage of the service area while maintaining reasonable distances between facilities.



(a) Adaptive gradient (b) Conditional gradient (c) Proximal gradient

Figure 2 Layout Comparison of Charging Base Stations Generated by the Three Algorithms

Figure 3 further compares the performance of each method across three key metrics: average distance to demand points, construction costs, and coverage rate. The results show that the adaptive gradient achieved the lowest average distance penalty of 7.80 units, representing a 25% improvement over the conditional gradient's 10.40 units. While its construction cost (3.21) was slightly higher than the proximal gradient's 3.03, it achieved the highest coverage rate of 65%, compared to 59% for both conditional and proximal gradients respectively. These results demonstrate that the adaptive gradient successfully balances the competing objectives of minimizing costs while maximizing service coverage.

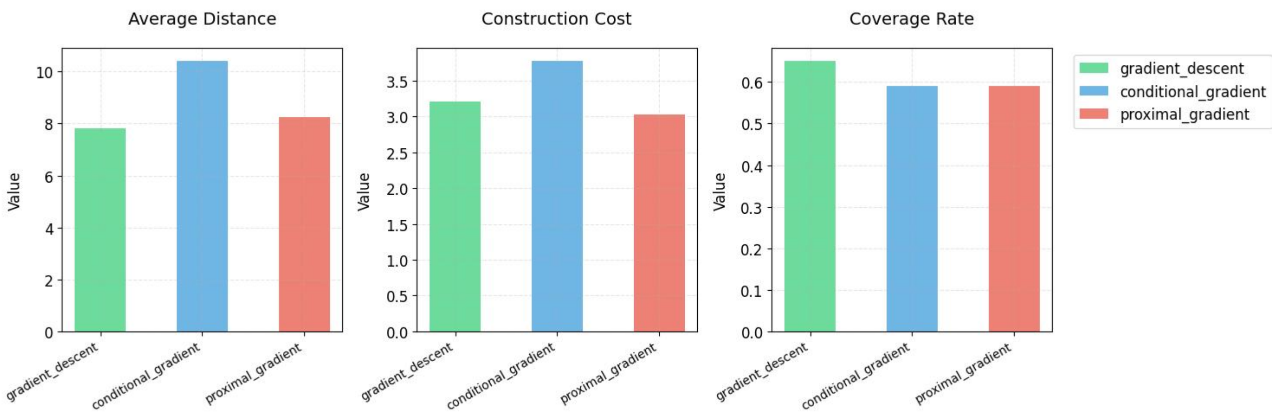


Figure 3 Cost Comparison of Charging Base Stations Generated by the Three Algorithms

In order to evaluate the scalability and robustness of the three optimization approaches, we conducted extensive experiments across four key problem dimensions: region size, demand point density, population center distribution, and charging station capacity. Each dimension was tested with four increasing scales to assess algorithmic performance under varying computational loads, as demonstrated in Table 1. Table 2 shows the mean and variance of results from five independent runs were reported.

Table 2 Performance Comparison of Different Algorithms in Varying Cases

Optimizer	Case (a)	Case (b)	Case (c)	Case (d)
Adaptive	5.26261± 0.732883	**9.938685± 0.905996**	**19.331808± 0.841872**	40.866991± 1.618788
Conditional	**5.213003± 0.748442**	10.623469± 0.214425	20.686607± 0.908285	41.086251± 1.783154
Proximal	5.290131± 0.703791	10.015033± 0.990372	19.625276± 1.278735	**40.536592± 1.609049**

The results revealed distinct performance patterns for each algorithm across different problem scales. The adaptive gradient demonstrates robust performance across medium to large-scale problems, while the conditional gradient excels in small-scale optimization scenarios. The proximal gradient becomes increasingly competitive as problem dimensions grow.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

CODE AVAILABILITY

The code to implement this project is available at https://colab.research.google.com/drive/1Ct4LfFBUKfm3Pmncx0OVWmOsO7RF4YqO?usp=sharing.

REFERENCE

[1] Zheng Y, Keith D R, Wang S, et al. Effects of Electric Vehicle Charging Stations on the Economic Vitality of Local Businesses. Nature Communications, 2024, 15(1): 7437.
[2] Du Z, Zheng L, Lin B. Influence of Charging Stations Accessibility on Charging Stations Utilization. Energy, 2024, 298: 131374.
[3] Mastoi M S, Zhuang S, Munir H M, et al. An In-Depth Analysis of Electric Vehicle Charging Station Infrastructure, Policy Implications, and Future Trends. Energy Reports, 2022, 8: 11504–11529.
[4] Rane N L, Achari A, Saha A, et al. An Integrated GIS, MIF, and TOPSIS Approach for Appraising Electric Vehicle Charging Station Suitability Zones in Mumbai, India. Sustainable Cities and Society, 2023, 97: 104717.
[5] Carra M, Maternini G, Barabino B. On Sustainable Positioning of Electric Vehicle Charging Stations in Cities: An Integrated Approach for the Selection of Indicators. Sustainable Cities and Society, 2022, 85: 104067.
[6] Kennedy J, Eberhart R. Particle Swarm Optimization//Proceedings of ICNN'95-International Conference on Neural Networks. IEEE, 1995: 1942–1948.
[7] Chakraborty N, Mondal A, Mondal S. Intelligent Charge Scheduling and Eco-Routing Mechanism for Electric Vehicles: A Multi-Objective Heuristic Approach. Sustainable Cities and Society, 2021, 69: 102820.
[8] Mason L, Baxter J, Bartlett P, et al. Boosting Algorithms as Gradient Descent//Advances in Neural Information Processing Systems. MIT Press, 1999.
[9] Grady S A, Hussaini M Y, Abdullah M M. Placement of Wind Turbines Using Genetic Algorithms. Renewable Energy, 2005, 30(2): 259–270.
[10] Braun G, Carderera A, Combettes C W, et al. Conditional Gradient Methods: arXiv:2211.14103. arXiv, 2023.
[11] Schmidt M, Roux N, Bach F. Convergence Rates of Inexact Proximal-Gradient Methods for Convex Optimization//Advances in Neural Information Processing Systems. Curran Associates, Inc., 2011.