International Journal of Science and Engineering Applications
Volume 14-Issue 01, 01 – 05, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1401.1001

Research on Text Detection of Electric Equipment Nameplates Based on DBNet

Tong Hu
School of Electronic Information and Electrical Engineering
Yangtze University
Jingzhou, China

Abstract: In the field of electrical equipment management, the accurate detection of text from nameplates is critical for effective information retrieval and maintenance. Traditional methods relying on manual detection are often time-consuming and error-prone. This paper presents a novel approach utilizing the DBNet (Differentiable Binarization Network) algorithm for automated text detection on electric equipment nameplates. DBNet employs an instance segmentation methodology, integrating differentiable binarization into the training process to produce robust binary maps, thereby enhancing text detection performance. Our experiments demonstrate that the DBNet model achieves superior accuracy and speed compared to traditional detection methods, effectively managing complex backgrounds and various text orientations. The results indicate that this approach significantly simplifies post-processing steps while maintaining high detection efficiency, making it a reliable solution for the challenges in the text detection of electric equipment nameplates.

Keywords: Computer Vision; Deep Learning; Electrical Equipment Nameplate; Text Detection;

1. INTRODUCTION

With the rapid development of artificial intelligence (AI) technology, its applications in the power industry have deepened, particularly in the fields of equipment management and maintenance. The stable operation of electrical equipment is directly related to the normal functioning of social production and daily life. However, traditional manual management methods are inefficient and prone to errors[1]. In recent years, Optical Character Recognition (OCR) technology has provided an efficient solution for the text detection of electrical equipment nameplates. OCR technology typically involves two key steps: text detection and text recognition. The former uses visual processing techniques to extract text instances from images, while the latter employs natural language processing techniques to retrieve the text content. These two steps are closely interconnected, with the accuracy of text detection directly influencing the final results of text recognition. Therefore, finding an efficient text detection method is of great significance.

To improve the efficiency and accuracy of electrical equipment information management, the introduction of AI technology for automated text detection of electrical equipment nameplates has become an important solution. By simulating the working mechanism of neural networks, deep learning models can automatically extract multi-level features and generate high-level abstract representations, making them particularly suitable for text detection in complex backgrounds. Compared to traditional methods, deep learning models not only achieve faster detection speeds but also accurately annotate text locations, significantly reducing the difficulty of subsequent recognition tasks.

Currently, the issue of data loss and low accuracy is prevalent in power system records, while nameplate information contains critical equipment parameters. The efficiency and precision of text detection are crucial for effective system management. Deep learning technology, with its powerful nonlinear approximation capabilities and feature learning advantages, can effectively handle the complex characteristics of nameplates, achieving precise mapping from images to text[2]. Based on this, this paper proposes a DBNet-based method for detecting electrical equipment nameplates. This method not only enhances the performance of text detection but also simplifies post-processing steps, significantly improving the accuracy and efficiency of nameplate text detection and providing reliable support for subsequent text recognition tasks.

2. DBNET ALGORITHM

With the rapid development of artificial intelligence technology, deep learning has been widely applied in various fields. The DBNet[2] algorithm proposes a scene text detection method based on instance segmentation, effectively addressing the challenges of detecting curved text. In the field of natural scene text detection, segmentation-based methods are favored due to their broad applicability. The typical workflow of these methods includes: first, outputting a probability map for text segmentation through a network; second, converting the probability map into a binary map using a set threshold; and finally, extracting detection results, namely the coordinates of the text boxes, through post-processing steps. However, a significant drawback of these methods is that the determination of the threshold is crucial. To address this issue, DBNet introduces the concept of differentiable binarization and integrates this step into the network training process, resulting in a more robust binary map and simplifying the post-processing workflow[4]. The DBNet network can detect both horizontal and multi-directional curved text, demonstrating superior performance while maintaining a fast detection speed compared to traditional text detection networks. Currently, researchers have proposed various improvements based on DBNet toenhance the accuracy of electrical equipment nameplate text detection.

International Journal of Science and Engineering Applications
Volume 14-Issue 01, 01 – 05, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1401.1001

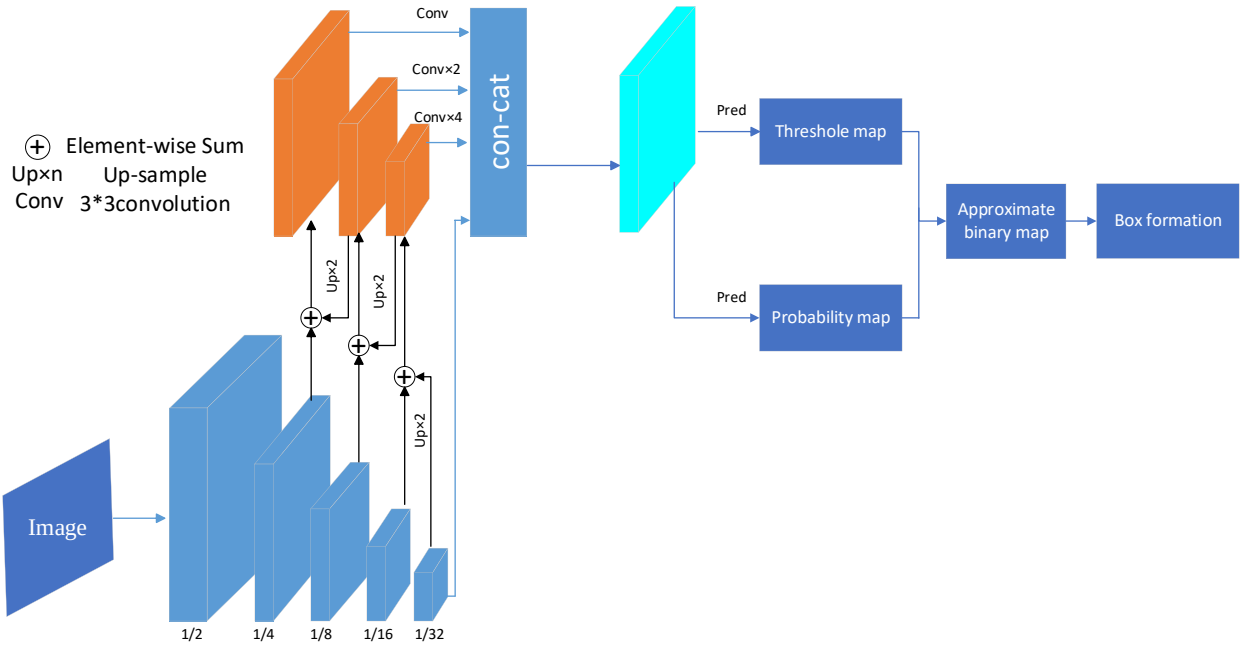


Figure. 1 DBNet Network Architecture

As shown in Figure 1, We can see that the DBNet architecture consists of three main components: the Backbone, the Neck, and the Head.

The Backbone is responsible for extracting features from the image. In this paper, we choose ResNet-50[5] as the backbone network, which is a relatively deep convolutional neural network containing 50 layers, along with feature submodules and convolution submodules. The feature submodules have the same dimensions for their input and output feature maps, allowing them to be concatenated. In contrast, the convolution submodules have different dimensions for their input and output feature maps, which prevents direct concatenation. Additionally, ResNet-50 employs a residual connection structure that facilitates direct connections across layers, helping to alleviate issues such as vanishing and exploding gradients, thereby allowing gradients to propagate more effectively and accelerating the model's convergence.

The Neck enhances feature extraction, and in this study, we utilize Feature Pyramid Networks (FPN) to achieve this goal. FPN can efficiently extract features at different scales from the image, which is particularly important for text detection. In text detection tasks, as the depth of the backbone network increases and the scale decreases, the final output feature maps tend to be capable of detecting only larger text while performing poorly on smaller text and edge text. FPN improves the detection capability for small target text and edge text by fusing shallow and deep features from the backbone network, providing stronger feature extraction for subsequent stages.

The Head is responsible for generating the probability map and threshold map for text regions, and through the differentiable binarization module, it generates the final binary map of text regions based on the probability and threshold maps.

2.1 FPN Structure

The FPN (Feature Pyramid Network) structure is designed to improve the feature extraction process by utilizing a multi-scale approach[6]. It creates a feature pyramid from a single input image, enabling the network to detect objects at various scales effectively. The FPN architecture typically consists of two main components: Bottom-Up Pathway and Top-Down Pathway.

The Bottom-Up Pathway builds a pyramid of feature maps by progressively downsampling the input image through several convolutional layers. Each layer captures features at different resolutions, with higher layers corresponding to lower spatial resolution but richer semantic information.

The Top-Down Pathway starts from the highest-level feature map and up-samples it to combine with the corresponding feature maps from the bottom-up pathway. This process enhances lower-level feature maps with more semantic information, improving the detection of smaller objects.

The FPN effectively integrates features from different levels of the backbone network, allowing for better detection performance across various object sizes, especially small and edge cases. This multi-scale feature extraction is crucial for tasks such as object detection and semantic segmentation.In the Feature Pyramid Network (FPN), feature maps from different layers are fused by addition. Through upsampling, feature maps of different scales are combined with feature maps that have undergone convolution operations. While this enhances the target feature information, it may also introduce background features. DBNet adopts a strategy of upsampling four groups of feature maps of different scales to the same scale for feature concatenation. This direct concatenation of raw features allows the network to learn how to effectively merge features, thereby reducing information loss. An example of the feature pyramid structure is shown in the figure below.

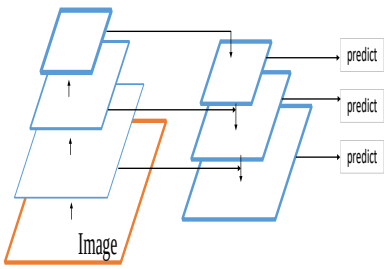


Figure. 2 FPN Architecture

International Journal of Science and Engineering Applications
Volume 14-Issue 01, 01 – 05, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1401.1001

2.2 Binarization

In digital image processing, one of the most common methods to extract target information is to set a threshold *T*. This threshold is used to divide the image into two parts: one part includes pixels with values greater than *T*, and the other part includes pixels with values less than T. This processing method is referred to as image binarization. Traditional threshold-based binarization methods are as follows:

$$B_{i,j} = \begin{cases} 1, & \text{if } P_{i,j} \geq t \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

Where *B* is the binary image; *i* and *j* represent the index positions in the features; *P* is the preceding feature map used to compute the binary image; and t is the threshold.The Sigmoid function is defined as follows:

$$F(x) = \frac{1}{1+e^{-x}} \quad (2)$$

The Sigmoid function is one of the most widely used non-linear activation functions in convolutional neural networks. It effectively maps input values to a range between 0 and 1, making it particularly suitable for models where outputs need to be interpreted as probabilities. However, during backpropagation, the Sigmoid function is prone to the vanishing gradient problem, which negatively affects the training of deep networks. Traditional Sigmoid-based binarization cannot be optimized during network learning. To address this issue, this paper introduces a Differentiable Binarization (DB) method, which serves as an approximate step function that can be effectively optimized during the network training process. The formula is as follows:

$$\hat{B}_{i,j} = \frac{1}{1+e^{-k(P_{i,j}-T_{i,j})}} \quad (3)$$

where *P* represents the approximate binary mapping; $T_{i,j}$ is the adaptive threshold mapping; and *k* is the learning factor. This approximate binarization function exhibits behavior similar to that of the standard binarization function, but it possesses the property of differentiability, allowing it to be optimized in synchronization with the segmentation network during training. DBNet enhances performance by optimizing the model through gradient backpropagation. For example, using binary cross-entropy loss effectively guides the network's learning process. Based on the aforementioned definition of the sigmoid function, we define the loss for the binary cross-entropy function as follows, where $x=P_{i,j}-T_{i,j}$ with the losses for positive and negative labels L^+ and L^- given by:

$$L_+ = -\log\frac{1}{1+e^{-kx}} \quad (4)$$

$$L_- = -\log(1-\frac{1}{1+e^{-kx}}) \quad (5)$$

2.3 Loss Founction

The DBNet text detection algorithm generates three core image outputs during model training: the probability map, the threshold map, and the binarization map. These images play a crucial role in the training process, especially when calculating the loss function, as they need to be compared with the corresponding ground truth labels to construct three different components of the loss function. The overall loss function is defined as follows:

$$L = L_s + \alpha \times L_b + \beta \times L_t \quad (6)$$

Where *L* is the total loss, L_s is the loss from the probability map, L_b is the loss from the binarization map, and L_t is the loss from the threshold map. α and β are weight coefficients. Both L_s and L_b use binary cross-entropy loss, defined as follows:

$$L_s = L_b = \sum_{i \in S_i} y_i \log x_i + (1-y_i)\log(1-x_i) \quad (7)$$

Here, S_i represents the sampled dataset with a positive-to-negative sample ratio of 1:3, x_i represents the predicted probability of being a text region, and *yi* represents the actual label value.

2.4 Evaluation Metrics

The evaluation metrics used in this study include Precision (P), Recall (R), and the F-measure (F). In a specific network model, after multiple training iterations, the values of Precision (P) and Recall (R) may fluctuate but generally remain within a certain range. To comprehensively assess the detection performance, this study introduces the F-measure, which is the harmonic mean of Precision and Recall, defined as:

$$P = \frac{TP}{TP+FP} \quad (8)$$

$$R = \frac{TP}{TP+FN} \quad (9)$$

$$F = \frac{2 \times P \times R}{P+R} \quad (10)$$

Where TP represents the number of correctly detected samples, FP represents the number of negative samples incorrectly detected as positive, and FN represents the number of positive samples incorrectly detected as negative. TP+FP is the total number of samples actually detected, and TP+FN is the total number of samples that should have been detected.

International Journal of Science and Engineering Applications
Volume 14-Issue 01, 01 – 05, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1401.1001

3. EXPERIMENT AND RESULT ANALYSIS

3.1 Model Building and Training

The DBNet model was established using segmentation-based methods in deep learning. The basic principles and steps of DBNet are as follows:

1) Data Preprocessing: Before performing text detection on the nameplates of power equipment, data preprocessing is necessary. This includes resizing images, cleaning the data, applying image augmentation, and removing non-standard annotation boxes to expand the training dataset size.

2) Feature Extraction: The input images pass through the Backbone network, undergoing a convolution and downsampling operation, resulting in four feature maps of different sizes.

3) Feature Enhancement: The extracted feature maps are fed into the Feature Pyramid Network (FPN) structure. After cascading through the FPN network, a feature map one-fourth the size of the original image is obtained.

4) Text Position Prediction: The head network predicts the probability map and threshold map using the cascaded feature maps. An approximate binary mapping is computed from the probability and threshold maps.

5) Model Training and Testing: DBNet is utilized to train and test the dataset, evaluating its intelligent detection capabilities for text detection on power equipment nameplates.

3.2 Experimental Results and Analysis

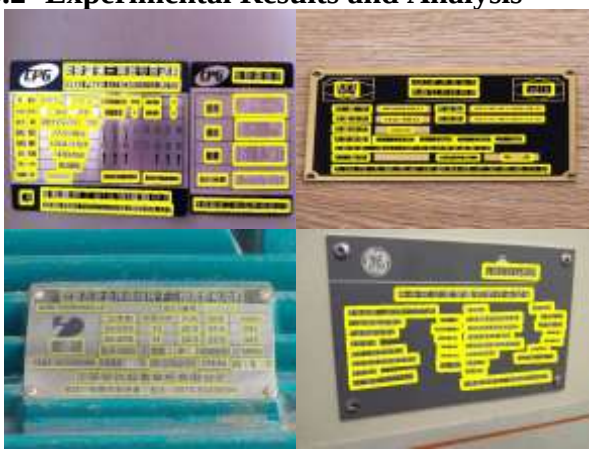


Figure.3 Detection Result

A comparison of the DBNet model with other models for text detection on power equipment nameplates is presented in Table 1.

From this table, it can be observed that DBNet outperforms other networks in overall performance due to its powerful feature extraction capabilities and the improved ability to generate threshold maps and binary maps at the detection stage. The DBNet-based text detection for power equipment nameplates excels in both accuracy and speed. In diversified scenario tests, this method accurately detects text regions and adapts well to complex backgrounds and embossed character features. Compared to traditional methods, DBNet shows significant improvements in detection efficiency and accuracy, providing reliable support for subsequent text recognition tasks.

4. CONCLUSION

In the field of text detection for power equipment nameplates, traditional methods primarily rely on manual efforts, which are time-consuming, labor-intensive, and prone to omissions, thereby causing inconvenience and risks during production or maintenance processes. To address these issues, this paper investigates text detection technology for power equipment nameplates based on deep learning. The DBNet algorithm is adopted to perform text detection on nameplates, and several attempts are made to improve detection performance.

DBNet is a segmentation-based text detection algorithm in deep learning, designed to detect text in images and annotate it in the form of bounding boxes. The algorithm introduces a differentiable binarization module, enabling the model to utilize an adaptive threshold map for binarization processing. This adaptive threshold map is incorporated into the loss calculation, which assists in optimizing the results during model training.

Experimental results show that this method not only significantly improves text detection performance but also simplifies the post-processing steps. Compared to other text detection models, DBNet demonstrates clear advantages in both effectiveness and performance. Generally, segmentation-based text detection methods require pixel-level prediction followed by post-processing algorithms to generate bounding boxes. However, post-processing algorithms are often complex and can lead to reduced computational speed. DBNet addresses this issue by integrating the binarization process into the training phase to enhance segmentation results, simplify post-processing, and maintain inference speed.

Table 1. Comparison Results

Models	P(%)	R(%)	F(%)
DBNet	80.2	72.4	76.1
PSNet	76.4	70.1	73.1
EAST	71.2	66.2	68.6

5. REFERENCES

[1] Wang Yifan, Wang Jiayu, Zhong Linlin, et al. Text Recognition of Electrical Equipment Nameplates Based on Deep Learning [J]. Electric Power Engineering Technology, 2022, 41(05): 210-218.

[2] Wang Jianxin, Wang Ziya, Tian Xuan. A Survey on Text Detection and Recognition in Natural Scenes Based on Deep Learning [J]. Journal of Software, 2020, 31(5): 1465-1496.

[3] Wang Daolei, Kang Bo, Zhu Rui. A Text Detection Method for Electrical Equipment Nameplates Based on Deep Learning [J]. Journal of Graphics, 2023, 44(04): 691-698.

[4] Liao M, Wan Z, Yao C, et al. Real-time scene text detection with differentiable binarization[C]//Proceedings of the AAAI conference on artificial intelligence. 2020, 34(07): 11474-11481.

[5] Zhang K, Sun M, Han T X, et al. Residual networks of residual networks: Multilevel residual networks[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2017, 28(6): 1303-1314.

[6] Xie J, Pang Y, Nie J, et al. Latent feature pyramid network for object detection[J]. IEEE Transactions on Multimedia, 2022, 25: 2153-2163.

International Journal of Science and Engineering Applications
Volume 14-Issue 01, 06 – 08, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1401.1002

Research on the Lightweight Path of Automotive Drum Brakes

Longjing Li
College of Automotive
Engineering
Zibo Vocational Institute
Zibo, China

Abstract: In the evolution of the automotive industry, the impetus from energy conservation and emission reduction policies has rendered automotive lightweighting an inescapable tendency. This article centers on dissecting the historical progression and forthcoming trends of structural lightweighting, lucidly expounding the essential significance of lightweighting across a multitude of sectors. Concurrently, it delves profoundly into the lightweight design methodologies for automotive drum brakes and also anticipates the future prospects of automotive drum brake lightweighting, thereby furnishing the automotive industry with a theoretical underpinning and practical counsel to fulfill the lightweighting objectives.

Keywords: Automotive Drum Brakes; Lightweighting; Future Outlook

1. INTRODUCTION

Amid the swift expansion of the automotive industry, the attainment of energy conservation, emission reduction, and performance enhancement has emerged as the central focus. Automotive lightweighting has thereby evolved into an inexorable trend. The automotive drum brake constitutes a vital component within the automotive braking framework. Its lightweighting holds substantial value in curtailing the overall vehicle weight, augmenting fuel economy, and bolstering braking efficacy[1].

2. THE EVOLUTION AND FUTURE TRENDS OF STRUCTURAL LIGHTWEIGHTING

2.1 Concept and Significance of Lightweighting

（1）Concept Explanation

Lightweighting refers to the reduction of material usage through design optimization and the employment of novel materials while guaranteeing product performance. The diminished material quantity can not only lower costs but also reduce resource consumption. Retaining strength and rigidity ensures the safety and reliability of products during utilization. Enhanced energy efficiency and performance enable products to satisfy the demands of contemporary technological progress.

（2）Explanation of Significance

Lightweighting is capable of efficiently diminishing energy consumption. In the context of automotive transportation, it leads to a reduction in energy usage, an enhancement of transportation efficiency, and a shortening of transportation time owing to the decreased load. Simultaneously, it contributes to environmental protection and sustainable development by lessening resource consumption and waste generation.

2.2 Development History

（1）Early Exploration Stage

Structural lightweighting initially found its application in the aerospace domain, aiming at cost reduction and performance enhancement. With the ceaseless advancement of composite materials and novel alloys, a foundation was laid for the actualization of lightweight design. Additionally, the design approaches based on shape optimization and topology optimization also propelled the progress of this development.

（2）Gradual Promotion Stage

In the automotive sector, the utilization of intelligent design and state-of-the-art materials has spurred the elevation of the automotive lightweighting level, consequently augmenting fuel economy and safety performance. Within the construction industry, by dint of new materials and technologies, structural stability has been ameliorated and construction efficiency has also been boosted. Additionally, diverse industries are perpetually probing and evolving high-performance materials, like composite materials and renewable materials, with the aim of attaining the objectives of environmental protection and high efficiency.

2.3 Analysis of the Current Situation

（1）Current Technologies and Methods

Typical lightweight materials encompass aluminum and carbon fiber, among others. Owing to their remarkable strength-to-weight ratios, they are extensively employed in areas such as aerospace and automotive industries. Contemporary design methodologies like topology optimization are capable of enhancing structural performance and minimizing material wastage. Computer simulation technology is utilized to perform virtual testing and analysis, thus augmenting the dependability of the design and optimizing cost-effectiveness[2].

（2）Market Demand and Challenges

The requirements for lightweighting technologies differ across diverse industries, making it essential to formulate customized and specialized solutions. Despite the fact that the continuous technological advancements can result in cost reductions, the initial investment is frequently substantial. Therefore, the payback period demands evaluation. In addition, matters concerning the recycling and disposal of products also merit attention to fulfill the objective of sustainable development.

2.4 Future Development Trends

（1）New Material Development

Renewable and eco-friendly materials will witness more extensive utilization to mitigate environmental pollution. Ultra-lightweight and high-strength materials continue to surface, presenting novel prospects for sectors such as aerospace and automotive. Nanotechnology holds vast application potential. It can enhance the performance of materials and be applied in a multitude of fields.

（2）Intelligent Manufacturing and Lightweighting

The combination of automated production lines and lightweight design is capable of enhancing production efficiency and decreasing costs. Artificial intelligence can refine the design and boost product quality and competitiveness. The analysis and feedback of real-time data contribute to the optimization of the production process.

（3）Summary and Prospect of Structural Lightweighting

Across diverse industries, data-driven decision-making, cross-border integration, and the elevation of employees' skills will all assume highly significant roles. Structural lightweighting exerts a notable promotional influence on environmental protection, resource recycling, and technological innovation. Concurrently, emerging technologies have also introduced entirely novel opportunities and possess more extensive application vistas in sectors such as healthcare, education, and smart cities..

3. RESEARCH ON THE LIGHTWEIGHTING PATH OF AUTOMOTIVE DRUM BRAKES

Conventional automotive drum brakes predominantly employ cast iron and composite materials. However, they suffer from issues like relatively high weight and subpar friction performance. Given the progressively prominent trend of automotive lightweighting, probing into novel materials and advanced technologies to surmount the lightweighting hurdles and fulfill the requisites of safety, durability, and economy has emerged as the central undertaking within the industry.

3.1 Lightweight Design Methods

（1）Material Selection

Novel lightweight materials possess the merits of low density and high strength. For instance, the utilization of magnesium alloys and aluminum alloys in automotive components has manifested a highly conspicuous weight reduction effect. Composite materials attain the objective of lightweighting while guaranteeing strength. It is of utmost importance to carry out a comprehensive contemplation and assessment of material strength and other characteristics (such as heat resistance, corrosion resistance, etc.) during the material selection process.

（2）Weight Reduction Design Principles and Analysis Methods

By way of proper material selection and structural optimization, the weight of components can be lessened, thereby enhancing the overall performance and fuel efficiency of automobiles. Employ finite element analysis to perform structural stress inspection in order to confirm the safety of the design and the lightweighting outcome. Contrast the traditional structures with the innovative ones and select substitute materials that are lighter and possess higher strength.

3.2 ANSYS Analysis Application

（1）Result Evaluation

Leverage ANSYS to execute numerical simulations regarding stress and displacement for automotive drum brakes, thereby pinpointing the structural weak points and potential failure areas. Illustrate the optimization process and outcomes via practical examples, and assess the efficacy of the lightweight design by means of quantitative metrics such as the weight reduction percentage and the performance enhancement scope.

（2）Experimental verification

Carry out practical tests on samples using precision instruments and gather data. Compare the outcomes from simulations with the measured data to appraise the precision and dependability of the model. Examine the influence of the test environment and conditions on the results to guarantee the repeatability and validity of the tests. Furnish optimization proposals according to the feedback, and ascertain the iteration path of the design and the anticipated performance of the final product.

4. Future Prospects of the Lightweighting of Automotive Drum Brakes

4.1 Prospects for Research on New Materials

The advancement of materials science will fuel the evolution of novel composite and functional materials, augment product performance, and reinforce environmental protection. In tandem with the progression of science and technology and the elevation of industries, the market appetite for new materials persists in swelling. Governments across the globe have successively rolled out policies to bolster the research and development of new materials with the aim of elevating the caliber of independent innovation.

4.2 Digital and Intelligent Design

The employment of digital tools in material selection and structural simulation is capable of enhancing the efficiency and precision of design. The intelligent algorithms generated by integrating machine learning and big data analysis can streamline the design process and reach the objective of intelligent design. Constructing a closed-loop feedback system allows for real-time modifications to the design plan, thus enhancing the performance of products.

4.3 Automated manufacturing process

Employing robotics and automated apparatus to construct intelligent production lines is able to enhance production efficiency and curtail labor costs. With the assistance of Internet of Things technology, the surveillance and optimization of the production process can be accomplished. By incorporating the concept of lean production, expenses can be slashed, quality can be enhanced, and the production cycle can be abbreviated.

4.4 Market demand and competitive landscape.

Presently, contemporary consumers have a stronger preference for environmentally friendly and sustainable products, and their demands regarding automotive performance and safety are on the rise. Simultaneously, the

need for personalized customization is continuously expanding. The competition within the industry has become more and more intense, with various competitors constantly innovating technologies and adjusting marketing strategies. The trends of electrification and intelligence in the automotive industry are prominent, and policies concerning sustainable development are impelling the industry to progress in an environmentally friendly direction.

5. CONCLUSION

The lightweighting of automotive drum brakes represents a crucial trend in the automotive industry's evolution. Via research on structural lightweighting and investigations into the means of reducing the weight of drum brakes, we have grasped the essential importance of material selection, the enhancement of design methodologies, the application of analytical techniques, and experimental verification. In the days to come, the progress of novel materials, digital and intelligent design, automated manufacturing, and the management of market demands will jointly facilitate more sustainable advancements in the lightweighting of automotive drum brakes, attain the objective of sustainable development within the automotive industry, and augment the overall performance and market competitiveness of automobiles.

During the exploration and implementation phases of automotive drum brake lightweighting, all industry stakeholders must closely track technological trends, actively confront challenges, and continuously innovate to adapt to shifts in market demands and contribute to the green and efficient advancement of the automotive sector.

6. ACKNOWLEDGMENTS

The author sincerely thanks the College of Automotive Engineering, Zibo Vocational Institute for its strong support of this research.

7. REFERENCES

[1] Song, X. Tu, Y.M. Chen, J.L. and Yuan, D. Simulation and Experimental Study on Performance Optimization of a Light Truck Lightweight Brake[J]. Agricultural Machinery and Equipment, 2022, 53(22): 135-137.

[2] Wang, Q.G. Wang, K. Zhao, F. and Wang, S.S. The Application of New Materials in the Lightweight of Heavy-duty Trucks[J]. AUTOMOTIVE TECHNOLOGY, 2016, (06): 53-54.

International Journal of Science and Engineering Applications
Volume 14-Issue 01, 09 – 13, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1401.1003

An In-Depth Analysis of Modern Caching Strategies in Distributed Systems: Implementation Patterns and Performance Implications

Mahak Shah
Department of Computer Science
Columbia University
New York, United States

Akaash Vishal Hazarika
Department of Computer Science
North Carolina State University
Raleigh, United States

Abstract: In the architecture of contemporary distributed systems, caching serves as a vital optimization strategy. This study explores the theoretical foundations, implementation patterns, and performance implications of various caching methodologies. We analyze caching architectures, highlighting their influence on system performance, scalability, and reliability. By synthesizing industry practices with theoretical frameworks, this paper provides insight into the selection and implementation of optimal caching strategies. In addition, we introduce innovative evaluation metrics to assess caching effectiveness in distributed environments and present empirical evidence supporting specific caching patterns for diverse use cases.

Keywords: distributed systems, caching strategies, machine learning optimization, performance optimization

1. INTRODUCTION

Modern distributed systems face significant challenges in managing data access patterns while ensuring system responsiveness and reliability. Caching has evolved from simplistic memory management to sophisticated distributed architectures, directly impacting application performance and structure. Several factors have driven this evolution:

- The exponential growth of data volume and user concurrency.
- Increasing demand for real-time processing and reduced latency.
- Geographic distribution of systems and users.
- Complex consistency requirements in distributed environments.
- The necessity for optimized resource utilization

Effective caching strategies must navigate competing considerations, including data consistency, operational complexity, and overhead. This paper aims to provide a comprehensive understanding of caching strategies to enhance performance in modern distributed systems.

2. BACKGROUND

We describe here the evolution of caching systems, performance metrics and some of the consistency models used while caching data

2.1 EVOLUTION OF CACHING SYSTEMS

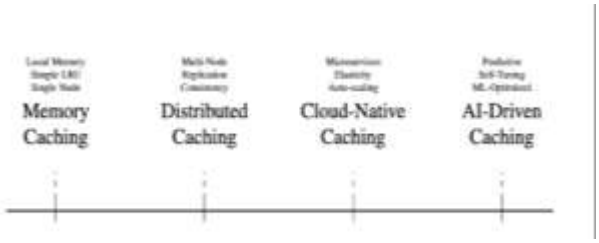


Figure 1: Evolution of Caching Systems

The evolution of caching systems[Figure 1] can be categorized into distinct eras.

The 1990s introduced Memory Caching, featuring local memory, simple LRU algorithms, and single-node deployments. The 2000s saw the emergence of Distributed Caching, marked by multi-node architectures, data replication, and consistency management. Cloud-Native Caching emerged in the 2010s, bringing microservices architecture and auto-scaling capabilities. Finally, the 2020s ushered in AI-Driven Caching, incorporating predictive analytics, self-tuning mechanisms, and ML-optimized systems.[Figure1]

2.2 Performance Metrics and Evaluation Framework

To evaluate the effectiveness of the caching system we propose a comprehensive framework that considers several dimensions of performance.

$$E = \frac{\alpha H + \beta L + \gamma C}{\delta R + \epsilon M}$$

Where

- E = Overall system efficiency
- H = Hit ratio (percentage of cache hits)
- L = Latency reduction factor
- C = Consistency measure
- R = Resource utilization
- M = Maintenance overhead
- The remaining greek variables are the weighing factors

2.3 Consistency Models

Model	Description	Use Case
Strong	Immediate consistency across nodes	Financial Transactions
Eventual	Allows temporary	Social Media

International Journal of Science and Engineering Applications
Volume 14-Issue 01, 09 – 13, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1401.1003

	inconsistencies	
Casual	Preserves cause-effect relationships	Messaging Systems

Table 1: Cache Consistency Models

Caching systems utilize various consistency models [Figure1] to maintain data coherence as shown in Table1.

3. CACHE ARCHITECTURE PATTERNS

3.1 Locality-Based Patterns

Cache patterns can be structured based on the relationship between data storage and data consumers, ranging from local caches that prioritize proximity to distributed caches that favor scalability. The choice of pattern significantly impacts system latency, network utilization, and overall application performance. These patterns represent different trade-offs between data proximity and system scalability[3]

Local Cache Implementation

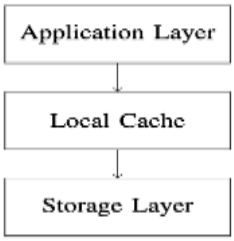


Figure2: Local Cache Architecture

The local cache [Figure2] implementation features three distinct layers:

- Application Layer: Primary interface for data requests
- Local Cache: Fast access memory storage
- Storage Layer: Persistent data storage.

Distributed Cache Architecture

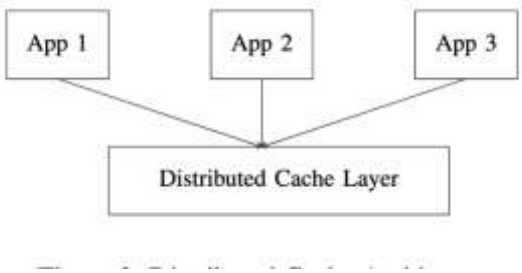


Figure3:Distributed Cache Architecture

The distributed cache architecture consists of:

- Multiple application nodes: app1, app2, app3
- Shared distribution cache layer
- Coordination mechanism for cache coherence.

3.2 Write Pattern Implementation

Synchronous Write Operation

Synchronous write operations ensure strong data consistency by updating both the cache and the underlying data store atomically. Although this approach introduces higher latency, it guarantees that cached data always reflect the state of persistent storage.

Key Characteristics include:

- Atomic updates to both the cache and the database
- Increased write latency
- Ensured transactional integrity and automatic rollback in case of failures.

Algorithm 1 Synchronous Write Pattern

```
1:  procedure SYNCHRONOUSWRITE(data, key)
2:      beginTransaction()
3:      if success then
4:          updateCache(key, data)
5:          updateDatabase(key, data)
6:          commitTransaction()
7:          return success
8:      else
9:          rollbackTransaction()
10:         return error
11:     end if
12: end procedure
```

Asynchronous Write Operation

Asynchronous write operations prioritize efficiency by decoupling cache updates from database updates. This approach is particularly beneficial for high-throughput situations where temporary inconsistencies can be tolerated.

International Journal of Science and Engineering Applications
Volume 14-Issue 01, 09 – 13, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1401.1003

Key Characteristics include:

- Immediate updates to cache with background database synchronization
- Reduced write latency
- Eventual consistency model
- Potential for temporary data inconsistencies

Algorithm 2 Asynchronous Write Pattern

```
procedure ASYNCHRONOUSWRITE(data, key)
    updateCache(key, data)
    success = queueBackgroundUpdate(key, data)
    if success then
        return true
    else
        invalidateCache(key)
        return false
    end if
end procedure
```

4. ADVANCED CACHING TECHNIQUES

4.1 Predictive Caching

Predictive caching uses machine learning models to predict data access patterns, potentially preloading data into the cache based on user behavior. This strategy aims to improve system performance by anticipating which data will be requested in the near future.

Big Data processing platforms like Spark uses this through lazy computation [5][6]

Key Concepts:

Machine Learning Models

Predictive caching algorithms commonly rely on machine learning techniques to analyze historical data access patterns. These models can identify trends in how users interact with the system, allowing for more intelligent predictions about future requests.

User Behaviour

By studying user interactions with the system, the predictive caching system can take into account various factors such as:

- Time of Day (e.g: users might request different data based on time)
- User roles (e.g: different roles accessing different datasets
- Recency of access (e.g: data that was recently accessed is likely to be requested again)
- Data relationships (e.g certain data is often accessed together)

Mathematical Representation

The Bayesian probability equation provides a framework for making predictions about data access based on contextual information:

$$P(access|context) = \frac{P(context|access) \cdot P(access)}{P(context)} \quad (2)$$

Where:

- P(access|context): The posterior probability, representing the probability of accessing a certain piece of data given the current context
- P(context|access): The likelihood, indicating how likely it is to observe a given context if a specific data item is accessed
- P(context): The evidence or the probability of the current context, serving as a normalization factor

4.2 Cache Replacement Policies

Modern cache replacement algorithms assess multiple factors using the scoring equation:

$$Score_{item} = w_1 F + w_2 R + w_3 S + w_4 C \quad (3)$$

Where:
• F = Frequency of access
• R = Recency of access
• S = Size of item
• C = Cost of retrieval
• w1, w2, w3, w4 = Weighting factors
The following replacement strategies are commonly implemented

LRU Cache

This strategy evicts the least recently accessed item when the cache is full. The underlying assumption is that data used will likely be used again soon. LRU maintains a list of items ordered by their access times to facilitate quick lookups

LFU Cache

LFU replaces the items that have been accessed the least often. It maintains a frequency count for each cached item, which can be updated upon every access. LFU is particularly effective when certain items are consistently accessed more than others.

FIFO Cache

This straightforward strategy removes the oldest item in the cache, assuming that older items are less likely to be used in the future. While simple to implement, FIFO does not consider usage frequency or recency, which can lead to suboptimal results.

International Journal of Science and Engineering Applications
Volume 14-Issue 01, 09 – 13, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1401.1003

Weighted Least Recently Used (WLRU)
An extension of LRU that assigns different weights to items based on their importance or usage characteristics. This strategy can outperform standard LRU in scenarios where certain items require more priority over others.

Random Replacement (RR)
In this approach, the item to be removed is chosen at random. While it may perform poorly in some situations, it is simple to implement and can occasionally be effective when access patterns are unpredictable.

Adaptive Replacement Cache (ARC)
ARC [4] dynamically adjusts its replacement strategy between LRU and LFU, maintaining two separate lists for each strategy. It balances recency and frequency based decisions, making it more versatile in various workloads.

5. IMPLEMENTATION CONSIDERATIONS

5.1 Technical Factors

Technical Factors	Considerations
Memory Usage	Balancing RAM allocation with dataset size
Network Latency	Effects of geographical distribution
Consistency	Aligning with business rules and SLAs
Access Patterns	Optimizing read/write ratios
Data Volatility	Characterizing update frequency

5.2 Operational Challenges

The operational challenges associated with caching strategies include:

• Ensuring cache coherence across distributed systems
• Managing network partitions and implementing effective recovery strategies
• Establishing monitoring and observability features for system performance
• Planning for capacity and scalability in response to fluctuating workloads
• Developing robust failure recovery and data restoration protocols

6. FUTURE DIRECTION AND CONSIDERATIONS

6.1 INNOVATIONS IN SERVERLESS PLATFORMS

The future of caching platforms shows promise for significant innovation across multiple dimensions. We anticipate enhanced flexibility in deployment options that will allow organizations to better customize their caching solutions. Support for various programming languages is expected to expand, making caching solutions more accessible to diverse development teams. Advanced local development and testing tools [7] will streamline the development process, while improved integration with cloud services will create more seamless deployments.

6.2 HYBRID ARCHITECTURES

As caching systems continue to mature, organizations are likely to gravitate toward hybrid architectures that offer greater versatility and optimization potential. These architectures will enable organizations to combine different caching strategies tailored to their specific needs, optimize for varying workload characteristics, and achieve a better balance between performance and cost considerations. The flexibility in deployment options will allow organizations to adapt their caching infrastructure as requirements evolve.

6.3 INDUSTRY STANDARDIZATION

Industry standardization efforts are expected to play a crucial role in shaping the future of caching systems. The development of unified protocols for cache interactions will facilitate better interoperability between different caching solutions. Standardized monitoring and metrics will enable more consistent performance evaluation and optimization. Common interfaces for cache implementations will reduce vendor lock-in, while portable configuration formats will simplify system management and migration processes.

6.4 AI AND MACHINE LEARNING INTEGRATION

The integration of artificial intelligence [8] and machine learning technologies [9] [10] promises to revolutionize caching systems. These technologies will enable improved prediction of access patterns, leading to more efficient cache utilization. Automated optimization of cache parameters will reduce manual configuration efforts and improve system performance. Intelligent resource allocation will enhance system efficiency, while advanced anomaly detection capabilities will help maintain system reliability and performance

7. CONCLUSION

In this paper, we have examined the evolution, implementation patterns, and performance implications of modern caching strategies in distributed systems. As data volumes and user expectations continue to escalate, effective caching mechanisms will be paramount. By balancing considerations such as consistency, latency, and system complexity, distributed systems can optimize performance and scalability.

REFERENCES

[1] M. Brown, "Evolution of Caching Strategies in Modern Distributed Systems," Journal of Systems Architecture, vol. 115, pp. 102-116, 2023.

[2] K. Davis and P. Wilson, "Consistency Models in Distributed Caching Systems," ACM Transactions on Database Systems, vol. 46, no. 3, pp. 1-28, 2023.

[3] J. Smith and B. Johnson, "Performance Analysis of Distributed Caching Architectures," ACM Computing Surveys, vol. 54, no. 2, pp. 1-34, 2022.

[4] X. Chen et al., "Adaptive Caching Strategies for Cloud Systems," IEEE Transactions on Cloud Computing, vol. 8, no. 4, pp. 1052-1065, 2023.

[5] A. V. Hazarika, G. J. S. R. Ram, and E. Jain, "Performance comparison of Hadoop and Spark Engine," in Proceedings of the 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, India, 2017, pp. 671-674.

[6] A. V. Hazarika, G. J. S. R. Ram, E. Jain, D. Sushma, and Anju, "Cluster analysis of Delhi crimes using different distance metrics," in Proceedings of the 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), Chennai, India, 2017, pp. 565-568.

[7] A. Chatterjee et al., "CTAF: Centralized Test Automation Framework for Multiple Remote Devices Using XMPP," in Proceedings of the 2018 15th IEEE India Council International Conference (INDICON), IEEE, 2018.

[8] R. Williams et al., "Machine Learning Approaches to Cache Optimization," in Proceedings of the International Conference on Distributed Computing Systems(ICDCS), pp. 245-254, 2023.

[9] Akaash Vishal Hazarika, Mahak Shah, "Serverless Architectures: Implications for Distributed System Design and Implementation," in International Journal of Science and Research (IJSR), vol. 13, no. 12, pp. 1250-1253, 2024.

[10] Anju, Hazarika A.V., "Extreme Gradient Boosting using Squared Logistics Loss function," in International Journal of Scientific Development and Research, vol. 2, no.8, pp. 54-61, 2017.

International Journal of Science and Engineering Applications
Volume 14-Issue 01, 14 – 16, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1401.1004

Research on the Application of ANSYS in the Optimization and Lightweight of Mechanical Structures

Longjing Li
College of Automotive
Engineering
Zibo Vocational Institute
Zibo, China

Abstract: This paper studies the application of ANSYS in the optimization and lightweight of mechanical structures. It introduces that ANSYS software can simulate a variety of physical phenomena and the functions of each analysis module, and elaborates on the importance of the optimization and lightweight of mechanical structures. It also explains in detail the application process in structural optimization and the lightweight analysis methods such as topological optimization and material replacement. The conclusion points out that ANSYS plays a crucial role and is constantly developing. In practical applications, its advantages should be fully exploited to promote the progress and sustainable development of the industry.

Keywords: ANSYS; Optimization of Mechanical Structures; Lightweight of Mechanical Structures

1. INTRODUCTION

In the field of modern mechanical engineering, with the continuous improvement of product performance requirements and the enhancement of environmental protection and energy-saving awareness, the optimization and lightweight of mechanical structures have become crucial research and development directions. ANSYS, a behemoth in the domain of engineering simulation software, wields unparalleled influence in this complex process. Its suite is replete with a vast array of analysis modules, each tailored to address specific engineering conundrums. For instance, the structural mechanics module can accurately simulate how a mechanical part will deform under different loads, be it tensile, compressive, or torsional forces. The fluid dynamics module, on the other hand, enables engineers to analyze the flow patterns around components, which is essential for optimizing heat dissipation or reducing aerodynamic drag.

Moreover, ANSYS's functionality extends far beyond basic simulations. It has advanced optimization algorithms that can iteratively modify design parameters. By setting clear performance goals, such as maximizing strength - to - weight ratio or minimizing energy loss, engineers can rely on ANSYS to sift through countless design alternatives. This not only aids in realizing the optimized design of mechanical structures but also has a domino effect on product performance. Improved designs lead to products that run more smoothly, with fewer breakdowns and longer service lives.

Cost reduction is another significant benefit. Through ANSYS - driven optimization, manufacturers can identify the most cost - effective materials and manufacturing processes. They can eliminate over - engineered components, reducing material waste and production time. Finally, in terms of meeting environmental requirements, lightweight structures designed with ANSYS consume less energy during operation, emit fewer pollutants, and are more conducive to recycling at the end of their life cycles, thus perfectly aligning with the pressing needs of environmental protection.

2. Overview of ANSYS Software

2.1 Software Introduction

ANSYS is a cutting-edge software that specifically focuses on engineering simulation, boasting an unrivaled reputation within the global engineering community. This powerful tool is equipped with highly sophisticated algorithms and a comprehensive set of computational models, enabling it to accurately simulate a diverse range of intricate physical phenomena. Whether it's the mechanical behavior of structures under extreme loads, the complex heat transfer processes in high-temperature environments, or the dynamic flow characteristics of various fluids, ANSYS can handle them with remarkable precision.

In the aerospace field, where safety and performance are of the utmost importance, ANSYS plays an indispensable role. Engineers rely on it to simulate the aerodynamic forces acting on aircraft wings during flight, optimizing their shapes to minimize drag and enhance lift. It also helps in analyzing the structural integrity of space vehicles, which must withstand the harsh conditions of space, including extreme temperature variations and intense gravitational forces. By accurately predicting potential weaknesses or areas of improvement early in the design phase, ANSYS significantly shortens the typically long and costly aerospace development cycles.

The automotive industry is another major beneficiary of ANSYS. With the ever-growing demand for fuel-efficient, high-performance, and environmentally friendly vehicles, car manufacturers turn to this software to simulate engine cooling systems. Through detailed fluid dynamics simulations, they can ensure that engines maintain an optimal operating temperature, improving fuel efficiency and reducing emissions. Moreover, in vehicle crash simulations, ANSYS can precisely model how different parts of a car deform upon impact, guiding the design of safer car bodies and reducing the need for expensive physical crash tests.

In the fast-evolving electronics sector, ANSYS is equally crucial. As electronic devices become smaller and more powerful, heat dissipation has become a major challenge. ANSYS's heat simulation capabilities allow engineers to analyze how heat is generated and dispersed within tiny chips and circuit boards, enabling them to design more efficient

cooling mechanisms. Additionally, it can simulate electromagnetic fields around electronic components, ensuring proper signal transmission and minimizing interference.

Overall, the high-precision simulation results provided by ANSYS have a profound impact on product development across these industries. By eliminating the need for numerous trial-and-error experiments, it effectively shortens the development cycle, saving both time and valuable resources. This reduction in development time directly translates into cost savings, as fewer prototypes need to be built and tested. Moreover, the optimized designs achieved through ANSYS simulations invariably lead to products with enhanced performance, better reliability, and increased competitiveness in the market.

2.2 Analysis Modules

The structural analysis module, a cornerstone within the ANSYS software suite, is specifically engineered to evaluate the fundamental mechanical properties of structures with a high degree of precision. It delves deep into assessing the strength of structures, determining the maximum load they can withstand without succumbing to failure, whether it's due to tensile, compressive, or shear forces. When it comes to stiffness, this module meticulously calculates how much a structure will deform under an applied load, which is crucial for applications where minimal deflection is required, such as in high-precision machinery or the frames of advanced optical instruments. Moreover, stability analysis is another key aspect; it predicts whether a structure will remain upright and functional under various loading conditions, safeguarding against catastrophic collapses, especially in large-scale construction projects like skyscrapers or long-span bridges.

The thermal analysis module, on the other hand, offers a comprehensive exploration of how materials behave when exposed to thermal environments. It takes into account a vast array of factors, starting from the basic thermal conductivity of materials, which dictates how quickly heat spreads through them. By simulating different temperature gradients, it can analyze how materials expand or contract, a phenomenon that could lead to misalignments or even fractures in tightly assembled components. Additionally, it studies the phase transitions that some materials undergo at specific temperatures, like the melting or solidification of metals, which can have a profound impact on the overall performance of a mechanical device. This in-depth understanding of thermal behavior provides designers with essential insights to optimize heat dissipation systems, prevent overheating, and ensure consistent operation across different temperature ranges.

The dynamic analysis module focuses on the complex responses of structures to dynamic loads. These dynamic loads can range from the rhythmic vibrations induced by rotating machinery, such as engines or turbines, to the sudden impacts experienced during earthquakes or collisions. By leveraging advanced algorithms, this module can simulate the time-dependent behavior of structures, tracking how they oscillate, resonate, or dampen vibrations over time. It enables engineers to predict potential fatigue failures that may occur due to repeated cyclic loading, which is a common issue in transportation vehicles like airplanes and trains. Moreover, understanding the dynamic responses helps in fine-tuning the design of shock absorbers, vibration isolators, and other damping devices, enhancing the overall durability and safety of mechanical systems.

Each of these meticulously designed and highly specialized modules, with their unique capabilities and detailed analytical outputs, provides indispensable and crucial bases for mechanical design. They empower designers to create more efficient, reliable, and optimized mechanical products, reducing the risks associated with trial-and-error approaches and accelerating the development process from concept to a fully functional and market-ready design.

3. The Importance of Optimization and Lightweight of Mechanical Structures

3.1 The Necessity of Structural Optimization

Optimization plays an absolutely pivotal role in the field of mechanical engineering. Through meticulous and scientific optimization processes, mechanical performance can be significantly enhanced. For instance, optimizing the structure of a machine part can refine its stress distribution, enabling it to withstand heavier loads and operate more stably under high-intensity working conditions. This not only improves the overall efficiency of the machinery but also reduces the likelihood of malfunctions.

When it comes to energy consumption, optimization is the key to achieving remarkable savings. By fine-tuning the design of mechanical systems, such as optimizing the gear ratios in a transmission system or streamlining the flow channels in a hydraulic device, unnecessary energy losses during operation can be minimized. This directly leads to a reduction in the long-term energy consumption of the product, which is highly beneficial in the context of today's increasing focus on energy conservation.

In addition, optimization is a powerful means of curbing material waste. With advanced simulation and analysis techniques, engineers can accurately determine the required amount of materials for each component, eliminating the overuse of materials that often occurs in traditional design methods. This not only saves valuable resources but also has a positive environmental impact.

The service life of products can be substantially prolonged through optimization. By improving the wear resistance of key components, optimizing the heat dissipation mechanism to prevent overheating-induced damage, and enhancing the corrosion resistance of metal parts, products can endure more usage cycles and maintain good performance over a longer period.

Cost reduction is another significant outcome of optimization. Lower energy consumption, less material waste, and fewer potential breakdowns all contribute to cutting down production costs, maintenance expenses, and even the cost of post-sales service. This cost advantage, in turn, strengthens the market competitiveness of products. Manufacturers can offer more competitive prices, attract more customers, and carve out a larger market share.

Moreover, optimized products are more agile in quickly adapting to market changes. In a rapidly evolving market where consumer preferences and technological advancements shift constantly, products that have undergone optimization can be more easily modified or upgraded. This flexibility allows businesses to stay ahead of the curve, responding promptly to emerging demands and trends, ensuring that their offerings remain relevant and appealing in the marketplace..

3.2 The Importance of Lightweight

Lightweight design has emerged as a crucial concept across various industries, bringing with it a multitude of benefits. Firstly, it significantly saves materials. In traditional manufacturing, excessive use of raw materials not only drives up costs but also places a heavier burden on natural resources. With lightweight design, engineers can precisely calculate the amount of material required for each component, eliminating redundant mass. This is especially evident in the automotive and aerospace sectors, where every gram of weight saved can translate into substantial savings over large production volumes.

Moreover, lightweight design plays a pivotal role in improving energy efficiency. Heavier objects demand more energy to move, whether it's a vehicle accelerating on the road or an aircraft taking off into the sky. By reducing the weight of these machines through lightweight techniques, the energy required for propulsion is slashed. For instance, in electric vehicles, a lighter body allows the battery to power the vehicle for a longer range on a single charge, thus enhancing overall energy utilization. In industrial machinery, lighter components also mean less power consumption during operation, contributing to long-term energy savings.

Another major advantage of lightweight design is its ability to meet stringent environmental protection standards. As global awareness of environmental issues intensifies, industries are under increasing pressure to reduce their carbon footprint. Lightweighting helps in this regard by minimizing the energy consumption associated with production, transportation, and usage of products. Additionally, through means such as topological optimization, rational material distribution and structural optimization can be achieved. Topological optimization algorithms analyze the stress, strain, and load conditions of a structure to determine the most efficient layout of materials. This ensures that materials are placed exactly where they are most needed, further reducing waste and enhancing the environmental friendliness of the product. For example, in the design of a new aircraft wing, topological optimization can identify areas that can be made thinner or hollower without sacrificing structural integrity, leading to a lighter wing that consumes less fuel during flights and emits fewer pollutants.

4. The application process of ANSYS in structural optimization

(1) Establishment of the Initial Design Model: Firstly, it is necessary to create a preliminary design model through CAD software or manual drawing to clarify the functional requirements and technical specifications of the product. At this stage, engineers need to determine the approximate shape and size of the structure according to the usage scenarios and expected performance of the product. For example, when designing a new type of engine cylinder block, factors such as power output and heat dissipation requirements should be taken into account, and a three-dimensional model of the cylinder block should be initially drawn.

(2) Setting of Analysis Conditions: Determining the structural materials, loading conditions and boundary conditions is the key basis for the subsequent ANSYS analysis. The choice of materials will affect the mechanical properties and costs of the structure. Different materials possess different characteristics such as strength, stiffness and density. Loading conditions include static loads and dynamic loads. For instance, in the design of a robotic arm, the static gravitational load when it grasps objects and the inertial load during the movement process should be taken into account. Boundary conditions stipulate the situation of the structure in terms of supports, constraints and so on.

(3) Result Evaluation and Iteration: Analyzing the simulation results is a crucial part of the optimized design. After obtaining the results of stress, strain, displacement and other aspects of the structure through ANSYS analysis, engineers need to determine whether these results meet the design requirements. If they don't, it is necessary to optimize the design scheme and conduct model iteration when necessary.

5. Methods of Lightweight Analysis Using ANSYS

5.1 Topological Optimization

It improves performance by adjusting the distribution of materials. Its process includes setting conditions, solving problems and generating results. It is widely applied in many fields and can significantly reduce weight and increase efficiency.

5.2 Material Replacement Analysis

Select materials according to the application scenarios, taking factors such as strength and stiffness into consideration. Cases show that reasonable material replacement can enhance the lightweight effect.

6. Conclusion

All in all, ANSYS plays an irreplaceable and important role in the optimization and lightweight of mechanical structures. Through its powerful analysis modules and functions, it can assist engineers in realizing the optimized design of structures, improving product performance, reducing costs, and meeting various requirements such as environmental protection. Meanwhile, with the continuous development of new technologies and the constant changes in market demands, ANSYS is also constantly innovating and evolving. In the future, it will play an even more significant role in numerous fields like automobiles, aerospace, and renewable energy, providing powerful support for the development of mechanical engineering. In practical applications, we should keep exploring and innovating, give full play to the advantages of ANSYS, promote the progress of mechanical structure optimization and lightweight technologies, and achieve the sustainable development of the mechanical industry.

7. REFERENCES

[1] Song, J. W., Qiu, R., and Zhou, G.H. Lightweight Design of the Rear Body of the Composite Automotive Front Floor[J]. Mechanical Design and Manufacturing, 2023, (02): 200-205.

[2] Li, M. L., Wang, Y.T. Research on the Lightweight of the Frame of a Light-duty Commercial Vehicle under Multiple Performance Constraints[J]. Mechanical Design and Manufacturing, 2024, (07): 006.

International Journal of Science and Engineering Applications
Volume 14-Issue 01, 17 – 20, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1401.1005

Ostrich Behavior Recognition Based on Deep Learning

Yusheng Duan
School of Electric Information
and Electrical Engineering
Yangtze University
Jingzhou, China

Abstract: Ostrich behavior is a key sign of their development and health. Quickly and accurately identifying it is crucial for growth monitoring and disease prevention. Computer vision technology, being real-time and non-contact, is widely used in livestock behavior recognition. But current methods, mainly for common livestock in simple settings, have drawbacks.This paper presents a method for ostrich behavior recognition using YOLOv7-MG. It aims to boost recognition efficiency and precision. Images of ostrich behavior are gathered from actual farms to create a dataset. The MobileOne network replaces the backbone of YOLOv7 to cut down computation and model size. Also, a GAM module is added to improve feature extraction in complex situations.The proposed method does better than YOLOv7 and other cattle behavior recognition systems. It has a relatively small model memory footprint and can precisely identify ostrich behavior. This lays the groundwork for ostrich disease prevention and management.

Keywords: Image recognition; Deep learning; Ostrich Behavior Recognition; image processing; Intelligent Farming

1. INTRODUCTION

Nowadays, intelligent breeding of livestock and poultry has increasingly become a prominent research area. Ostriches, in particular, possess high economic value, not only due to their valuable products such as meat, eggs, and feathers which are in significant market demand [1], but also because of their rapid growth rate, relatively lower costs, and shorter breeding cycles. However, ostrich farming is not without its challenges. Disease prevention and control is a crucial aspect, as ineffective management can result in substantial financial losses. Traditional approaches like manual inspection and symptomatic preventive medicine have proven to be insufficient, being hampered by low efficacy, high costs, and a narrow monitoring range .

The emergence of artificial intelligence and computer vision technologies has brought new possibilities to the livestock and poultry breeding industry. In the domain of behavior recognition, several techniques have been explored. For instance, Nasirahmadi and colleagues employed deep learning techniques such as Faster R-CNN to identify the postures of pigs [2]. Liu et al. utilized an enhanced YOLO v3, optimizing it with the amplitude iterative pruning approach to achieve a 79.9% recognition accuracy for cow feeding behavior [3]. Kim et al. adopted the YOLOv3 and YOLOv4 models to identify the eating behavior of suckling pigs [4]. Wang et al. proposed an enhanced YOLOv5-based cow behavior detection system for accurately identifying cow behavior during estrus [5].

Nonetheless, these existing methods are mainly designed for relatively simple and less complex environments, and they often lack the adaptability required for the more intricate and dynamic conditions of ostrich farming. The free range of activity for ostriches is considerably larger compared to that of pigs and cows, with a greater likelihood of other objects being present in their activity area, thus making the living environment more complex.

Therefore, this paper centers around ostrich behavior recognition. Initially, data is meticulously gathered from the actual ostrich breeding environment to construct a comprehensive dataset. Subsequently, taking YOLOv7 as the foundation, the model is refined by substituting the YOLOv7 backbone network with the lightweight MobileOne backbone network, effectively reducing the number of parameters and computational load . Finally, the Global Attention Mechanism (GAM) module is incorporated into the header to augment the model's capacity to extract features from complex surroundings. The ultimate objective is to propose a lightweight and efficient ostrich behavior recognition model that can serve as a valuable reference for the intelligent management of ostrich farming, enhancing both productivity and the overall health and well-being of the ostrich population.

2. MATERIAL AND METHODS

2.1 Experimental Data Acquisition and Processing

The experimental dataset was collected from an ostrich farm. The area where the ostriches were naturally active was chosen as the data collection site to ensure the universality of the data. Mobile phones were then utilized to take pictures of the ostriches from various angles at a suitable distance outside the farm.

The video recorded contains common postures of ostriches. A third-party Python module was employed to extract one frame every ten frames and export the movie as JPG files. To avoid the overfitting problem caused by the high visual similarity of adjacent frames and the single ostrich feature, data cleaning was carried out. After manual inspection and screening, 800 photos were retained for further processing and experimentation. Subsequently, some images with relatively low quality and excessive similarity were removed.

2.2 Data Enhancement

In order to strengthen the robustness of the model, 800 negative samples were included in the ratio of positive and negative samples 1:1, totaling 1600 pictures as the original image dataset. It is quite likely to result in issues like an unstable training process and overfitting of the model if these 1600 photos are utilized straight for training and testing. This

International Journal of Science and Engineering Applications
Volume 14-Issue 01, 17 – 20, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1401.1005

work expanded the original dataset from 1600 to 6400 pictures using data augmentation techniques, such as varying image brightness and introducing Gaussian noise, to improve model performance and generalization capacity.

Figure. 2 Improved YOLOv5s

2.3 Algorithmic Modeling and its Improvement

Chien-Yao Wang and Alexey Bochkovskiy et al. created the YOLOv7 model in 2022. The model incorporates a number of techniques, such as model reparameterization, mod-el scaling based on tandem models [6], and E-ELAN (Extended Efficient Layer Aggregation Network) [7]. The backbone network, header network, and prediction network are the three primary parts of the YOLOv7 network. The MP module, which does preliminary feature extraction on the input pictures, the E-ELAN module, and the CBS module make up the majority of the backbone network.

The primary components of the head network are the MP, ELAN-H, Concat, and Spatial Pyramid Pooling and Convolutional Spatial Pyramid Pooling Sppcspc modules. These modules combine and improve the incoming characteristics of the backbone net-work to extract the most important information. After using the Rep structure to help train the prediction network, the number of image channels for the head network's output fea-tures is adjusted using 1x1 convolution, and the predicted correlation data is eventually acquired.The network structure of the YOLOv7-MG lightweight ostrich behavior recognition model proposed in this paper is shown in Figure. 2.

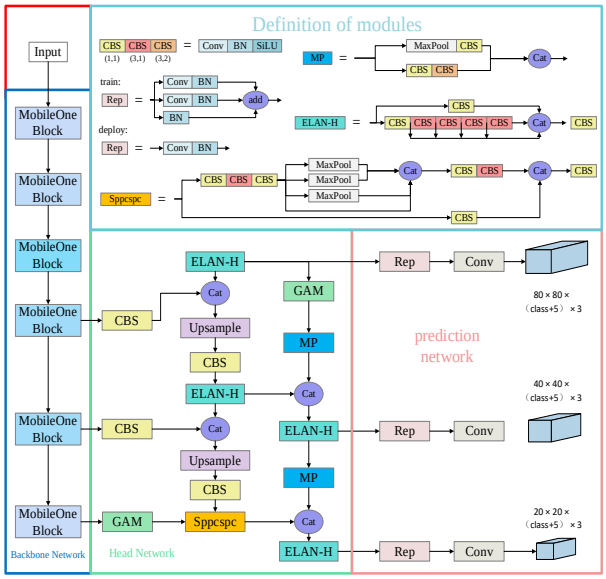


Figure. 2 YOLOv7-MG structure diagram

2.3.1 *Replacement of the backbone network*

The applicability and practical usefulness of the model application to the mobile terminal may be shown in the recognition of ostrich behavior and further study in this ar-ea. As a result, the model must have less complexity; in other words, it must be light-weight. In this study, we replace the original YOLOv7 backbone network with a light-weight one called MobileOne[8]. The original goal of MobileOne, a lightweight convolu-tional neural network, was to maximize model performance and compute economy in situations involving mobile or edge applications. Several MobileOne Blocks make up the MobileOne paradigm, and these MobileOne Blocks are interconnected. The step size and the number of output channels are the sole distinctions between the various MobileOne Blocks that make up the MobileOne paradigm. Depthwise Convolution + Pointwise Con-volution provides the fundamental structure of the MobileOne Block structure, which in-corporates the parameterization concept from RepVGG [9]. Fig. 2 depicts the MobileOne Block construction.

During training, the depth convolution section consists of k blocks of 3×3 depthwise convolution, one branch of 1×1 depthwise convolution, and one batch normalization (BN) branch in parallel. The point convolution section includes k blocks of 1×1 pointwise con-volution and one BN branch in parallel. In contrast, during inference, the MobileOne Block structure has no branches and follows a streamlined architecture. This design ensures ef-fective feature extraction during training while providing fast inference and low model memory usage during prediction.

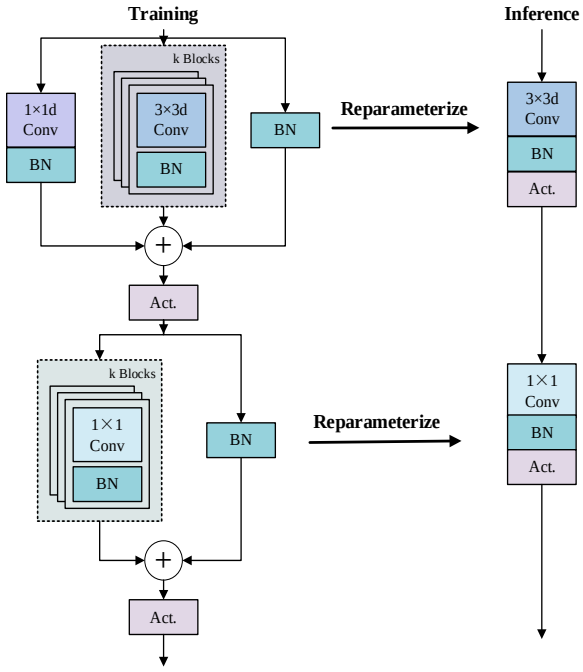


Figure. 3 MobileOne Block structure diagram

2.3.2 *Introduction of attention mechanisms*

It can be observed from the dataset photographs that ostriches provide shade to each other, as well as to themselves, due to the surrounding fences and trees. As a result, the complexity of the environment presents a greater challenge for the model's ability to ex-tract features. The attention mechanism helps the network model focus more on the im-portant aspects

of the image. The popular SENet[10] and CBAM [11] attention modules (Convolutional Block Attention Module

International Journal of Science and Engineering Applications
Volume 14-Issue 01, 17 – 20, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1401.1005

and Squeeze and Excitation Network, respectively) The connections between dimensions are weakened when Attention Modules overlook the interactions between space and channels. On the other hand, by improving the interaction of information in global dimensions, GAM [12] can lessen the dispersion of important in-formation in features and enhance the network's overall capacity for critical feature ex-traction. It is composed of a spatial attention sub-module (e.g., Fig. 6) and a channel atten-tion sub-module (e.g., Fig. 5), as seen in Fig. 4.

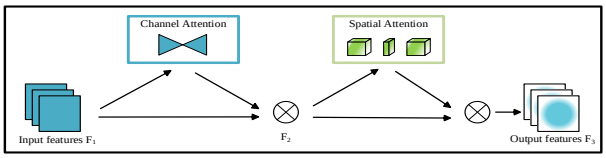


Figure. 4 GAM structure diagram

In Figure 4, the results of the GAM attention module, denoted as F_3, are obtained by multiplying the input feature map F_1 by the channel attention map $M_c(F_1)$ element-wise, followed by multiplication of the intermediate feature map F_2 with the spatial attention map $M_s(F_2)$ element-wise. The diagram clearly indicates these element-wise multiplication operations, highlighting the importance of each attention mechanism.

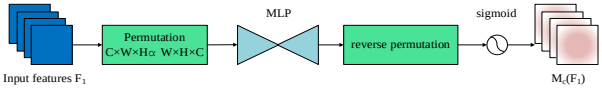


Figure. 5 Structure of the channel attention sub-module of the GAM module

As shown in Figure 5, the channel attention submodule first retains the information across three dimensions (channel, spatial width, and spatial height) by using 3D permu-tation. Then, it magnifies the cross-dimensional channel-spatial dependencies with a multi-layer perceptron (MLP). The MLP employs an encoder-decoder structure and re-duces the dimensionality through a reduction ratio to improve computational efficiency and performance.

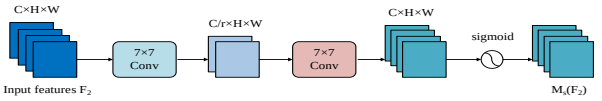


Figure. 6 Structure of the spatial attention sub-module of the GAM module

As shown in Figure 6, the channel attention submodule in the GAM module consists primarily of a 3D transformation operation, which allows it to preserve the spatial, chan-nel, and depth-related information of the input features. This transformation is followed by a two-layer Multi-Layer Perceptron (MLP), which enhances the dependencies between the channel and spatial dimensions. The spatial attention submodule, on the other hand, is composed of two 7x7 convolutional layers designed to efficiently fuse spatial infor-mation. In the first convolutional layer, the number of channels is reduced from C to C/r, where r is a reduction ratio hyperparameter that controls the degree of dimensionality re-duction to balance between computational efficiency and information retention. Together, these operations in the GAM module enable the network to focus on regions with signifi-cant contextual importance, while reducing the dispersion of meaningful feature infor-mation, thus improving the overall performance of the model.

3. EXPERIMENTAL SETUP AND RESULTS

3.1 Experimental Environment and Parameter Settings

The operating environment for this experiment features a 12th Gen Intel(R) Core(TM) i5-12400F CPU, 16GB of RAM, and an NVIDIA GeForce RTX 3060 GPU. The experiments are conducted within the PyTorch deep learning framework, with a CUDA version of 11.1.

During the training, the initial learning rate (Base_lr) is configured as 0.01, the weight decay is set to 0.0005, the optimizer (Optimizer) is chosen as SGD, and the batch size (Batchsize) is set to 8.

3.2 Experimental Results

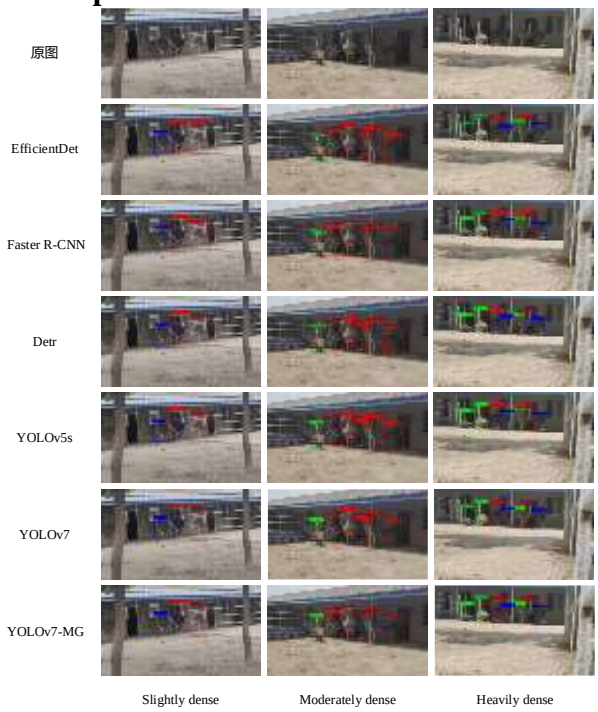


Figure. 7 Comparison of ostrich behavioral recognition results in environments with different densities

Figure. 7 shows a comparison of the results of ostrich behavior recognition in three different densely populated environments (slightly dense, moderately dense and heavily dense). Under the slightly dense environment, all the ostrich's behaviors were accurately identified. In the moderately dense environment, most of the ostrich behaviors were accurately identified, and the accuracy of individual ostrich behaviors that were not too badly occluded decreased slightly. In the heavily dense environment, two ostriches' behaviors were missed due to severe occlusion, and the rest of the ostrich behaviors were accurately identified. Each model was able to identify the target that should be identified, and there were differences in the level of confidence, accuracy of target frame localization, and detection speed between the different models. Although EfficientDet, FasterR-CNN, and DETR were slightly less accurate in localization, the overall performance was still quite good.YOLOv5s and YOLOv7 were balanced and efficient in terms of speed and accuracy.YOLOv7-MG was the best performer, with optimal localization and recognition accuracy, and was the fastest in terms of recognition speed. This indicates that the robustness

of the model in this study is high, with only a few misrecognitions in heavily dense environments.

4. CONCLUSION

In this study, an ostrich behavior recognition approach based on enhanced YOLOv7 is proposed to identify the daily behaviors of farmed ostriches. To enhance the model's efficiency and adaptability, the YOLOv7 backbone network is replaced with the lightweight MobileOne backbone network, reducing computational complexity and parameter numbers. Subsequently, a global attention mechanism (GAM) module is incorporated into the head network to improve the model's feature extraction capabilities in complex environments.

Compared to the original YOLOv7 model, the proposed method demonstrates notable improvements in various aspects. It achieves a significant enhancement in mean average accuracy and recognition speed, while also optimizing the model size. The experiments are conducted using a self-built dataset, as there is currently no relevant public dataset available.

When compared to other popular models such as EfficientDet, Faster R-CNN, Detr, YOLO5s, and the original YOLOv7, the approach presented in this research shows superiority in both identification speed and mean average accuracy. This indicates that the YOLOv7-MG model proposed in this paper outperforms existing models in recognizing ostrich behavior in complex environments, with better robustness.

However, it should be emphasized that the acquisition of high-quality datasets of ostrich behavior images remains a major challenge in the development of ostrich smart farming. Future efforts should focus on collecting a sufficient number of high-quality images of abnormal ostrich behaviors, which will lay the foundation for the advancement of ostrich abnormal behavior detection and the intelligent prevention and control of ostrich diseases.

5. REFERENCES

[1] Medina F X, Aguilar Moreno E. Ostrich meat: nutritional, breeding, and consumption aspects. The Case of Spain[J]. 2014.

[2] Nasirahmadi A, Sturm B, Edwards S, et al. Deep learning and machine vision approaches for posture detection of individual pigs[J]. Sensors, 2019, 19(17): 3738.

[3] Liu Yuefeng, Bian Haodong, He Yingjie, et al. A Recognition Method for Multi-target Cow Feeding Behavior Based on Amplitude Iterative Pruning [J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53 (02): 274 - 281.

[4] Kim M J, Choi Y H, Lee J, et al. A deep learning-based approach for feeding behavior recognition of weanling pigs[J]. Journal of animal science and technology, 2021, 63(6): 1453.

[5] Wang R, Gao Z, Li Q, et al. Detection method of cow estrus behavior in natural scenes based on improved YOLOv5[J]. Agriculture, 2022, 12(9): 1339.

[6] Gao P, Lu J, Li H, et al. Container: Context aggregation network[J]. arXiv preprint arXiv:2106.01401, 2021.

[7] Dollár P, Singh M, Girshick R. Fast and accurate model scaling[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 924-932.

[8] Anasosalu Vasu P K, Gabriel J, Zhu J, et al. An Improved One millisecond Mobile Backbone[J]. arXiv e-prints, 2022: arXiv: 2206.04040.

[9] Ding X, Zhang X, Ma N, et al. Repvgg: Making vgg-style convnets great again[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 13733-13742.

[10] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.

[11] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.

[12] Liu Y, Shao Z, Hoffmann N. Global attention mechanism: Retain information to enhance channel-spatial interactions[J]. arXiv preprint arXiv:2112.05561, 2021.

International Journal of Science and Engineering Applications
Volume 14-Issue 01, 21 – 24, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1401.1006

Design and Implementation of Transparent Liquid Concentration Measurement Based on ARM

Ni Qiu
Yangtze University
School of Electric Information and Electrical Engineering
Jingzhou, China

Abstract: At present, there have been many research results in liquid concentration detection, but China is still relatively backward in optical detection technology, and it is of great significance to research and develop new liquid detection systems. Under the same solute, due to the different concentration of the liquid, its absorption degree of fixed incident light will also be different, the use of photoelectric sensor to detect the change of light intensity, you can obtain the absorbance of the liquid, and then calculate the liquid concentration. Based on the above theory, this paper designs a liquid concentration measurement system based on the main control module of STM32F103RCT6 chip based on ARM Cortex-M3 core. The system consists of an STM32F103RCT6 chip, a TSL2591 photoelectric sensor, a laser emitter, and an OLED display. When the liquid concentration in the container changes, it will cause the light intensity change detected by TSL2591, and according to the change of the obtained data, the data will be fitted with the change of the configured liquid concentration, and finally a curve corresponding to the liquid concentration and the data change is obtained. The value measured by the TSL2591 is substituted into the curve to obtain the concentration value of the liquid and displayed on the OLED display. Through tests, it has been proved that the device can determine the concentration of liquids more accurately. The system has the advantages of simple structure, high sensitivity and fast response time, and can accurately measure the concentration of liquid.

Keywords: ARM Cortex-M3 core;TSL2591 light sensor;OLED screen

1. INTRODUCTION

In recent years, the rapid development of liquid detection technology, involving almost all aspects of the production process, in many production fields, not only requires the simple operation of measuring instruments, but also requires high accuracy of liquid measurement, such as in the production of drugs, the need for accurate configuration of the proportion of various medicinal materials and the concentration of various drugs in order to make successful drugs. Especially in recent years, the outbreak of the epidemic, the country's demand for drugs is increasing day by day, at this time, the detection of drug concentration is particularly important, so detection technology is one of the directions of people's research.

There are various methods for measuring liquid concentration, and after years of development, the detection technology has also developed relatively maturely, and the methods that have been applied and popularized include capacitance method, supergenerated grating method, grazing incidence method, etc.[1]. The design of this project is based on a novel optical method, which mainly uses photoelectric sensors to measure the concentration of liquids. The detection of liquid concentration by using the photoelectric sensor method has the characteristics of fast, simple, no contamination of the liquid to be measured, and easy signal conversion to achieve automatic control.

2. GENERRAL DESIGN

In order to complete this test task, the STM32F103RCT6 chip based on the ARM Cortex-M3 core is selected as the main control module of the measurement device, and the TSL2591 photoelectric sensor is also selected as the main test element, and the light intensity digital signal converted by the TSL2591 photoelectric sensor is received by the IIC communication principle of the STM32F103RCT6 chip. The data is displayed on the OLED screen via the IIC[2]data bus, i.e. the liquid concentration.

2.1 Experimental Principle

Lambert-Beale law: is the basic law of light absorption, which describes the relationship between the intensity of light absorption by a substance at a certain wavelength and the concentration of light-absorbing substances and the thickness of their liquid layer[3]. The relationship is shown in equation (1).

$$A=Kbc \quad (1)$$

Where K is the molar absorbance coefficient of the solute, b is the range of light passing through the solute, and c is the concentration of the liquid. From the basic idea of Beer's law, it can be seen that absorbance is directly proportional to the concentration of the liquid, that is, the greater the concentration of the liquid, the greater the absorbance. The absorbance is the intensity of the incident light before the light passes through the liquid and the transmitted light intensity after the light passes through the liquid, as shown in equation (2) (I_0 is the light intensity of the control group, I_1 is the light intensity of the liquid to be measured).

$$A=lg\frac{I_0}{I_1} \quad (2)$$

The basic principle of this experiment is that under the same solute, the concentration of the liquid will be different, and the degree of absorption of light will also be different. Therefore, the relationship between the light intensity of different concentrations of liquids and the light intensity of the same beam of light through different concentrations of liquids is studied, and the concentration of liquids can be accurately measured by using this relationship.

2.2 Modle Choice

2.2.1 Main Control Module

In this design, a STM32F103RCT6 microcontroller was selected as the main control module, and the chip is based on the ARM Cortex-M3 core, which can perform complex data

International Journal of Science and Engineering Applications
Volume 14-Issue 01, 21 – 24, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1401.1006

processing. The main frequency of the clock is relatively high, the maximum speed of the CPU can reach 72MHz, and at the same time, the chip integrates 12-bit precision ADC, USART serial port and other complex circuits[4].

2.2.2 Sensor Module

Sensor, also known as transducer, is a device that converts information into electronic information signals, usually composed of sensitive originals and conversion originals, which can convert the obtained information into electronic information signals in accordance with certain specifications and rules, so that the obtained information can meet the requirements of information transmission, storage, display, recording and management. In this paper, TSL2591 photoelectric sensors are mainly used for liquid detection. The photoelectric sensor module is TSL2591[5], which is an extremely sensitive optical-to-digital converter that converts light intensity into a digital signal and outputs it directly through the IIC interface.

2.2.3 Display Module

The display module uses a 0.96-inch (4-pin) OLED[6] display. The module has the characteristics of small size, high resolution, self-illumination, and a variety of interface methods.

2.2.4 Key Detection Module

The key detection module adopts an external interrupt mode, the STM32 microcontroller receives the digital signal sent by the photoelectric sensor, presses the button, and the STM32 microcontroller sends the signal to the screen, that is, the screen displays the concentration of the liquid to be measured.

2.3 Device Design

Figure 1 shows the basic module of the device, which consists of an STM32 main controller, a photoelectric sensor module, a laser emitter, a display module, and a button module. The laser emitter irradiates the liquid, and the photoelectric sensor module converts the light intensity signal into a digital signal and transmits it to the STM32 main controller.

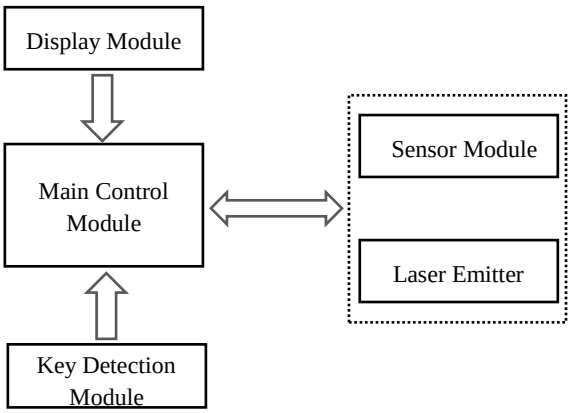


Figure. 1 The basic module of the device

The actual drawing of the device is shown in Figure 2.

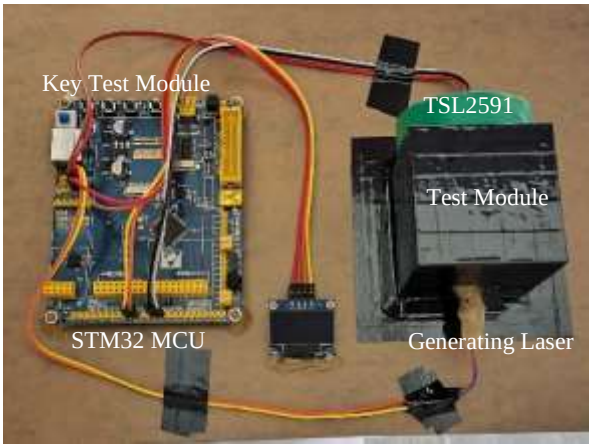


Figure. 2 Device diagram

3. DESIGN OF EXPERIMENTS

3.1 Liquid Selection

Two reagents were prepared for this design: saline and sugar water.

It is understood that the absorbance of salt water does not exceed 0.650, which is not suitable as a detection object, and the absorbance of sugar water is much greater than that of salt water, so the liquid selected for measurement in this design is sugar water, sugar water is easy to obtain, and the operation of configuring different concentrations is relatively simple, the larger the concentration of sugar water, the greater the density of the liquid, and the change of light intensity through sugar water also becomes larger, and the absorbance changes greatly, which is convenient for observation and analysis in the experiment. experiment.

3.2 Experimental Procedure

According to the principle of concentration detection, the experimental process design was completed.

(1) Configure 6%-14% sugar water liquid with a gradient concentration of 1% (due to the fixation of the experimental device, the actual operation is to add sugar to the low-concentration liquid to increase the concentration);

(2) Measure the intensity of the incident light emitted by the laser emitter through the clear water, wait for the device to be stabilized, read multiple sets of data, and take the average value as the control group of this experiment;

(3) Replace different concentrations of sugar water liquid into a square container in turn, read the data of each concentration of liquid for multiple times, take the average value, and find the transmitted light intensity after the light passes through different liquid concentrations;

(4) According to the transmitted light intensity and incident light intensity of the corresponding concentration, the absorbance of the corresponding concentration is calculated, and the specific mapping relationship between the liquid concentration and absorbance is fitted by MATLAB[7];

(5) The curve equation obtained by fitting the curve is written into the program through the STM32Cube, and then the program is burned into the single-chip microcomputer;

International Journal of Science and Engineering Applications
Volume 14-Issue 01, 21 – 24, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1401.1006

(6) Configure different concentrations of sugar water liquid, and test the error of the device to measure the concentration of liquid.

3.3 Fitting Phase

Before starting the test, the relationship between the absorbance of the liquid and the concentration of the liquid needs to be fitted to obtain its specific mapping. In the process of testing the data, it is necessary to ensure that the experimental environment remains unchanged, and the flow chart of the main program in the fitting stage is shown in Figure 3.

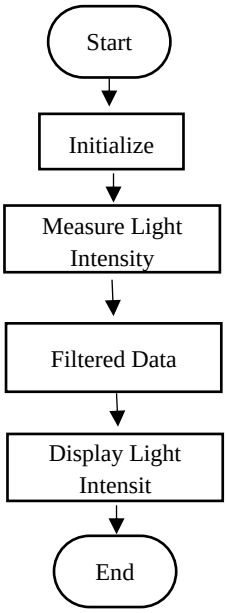


Figure. 3 Fitting phase process

After the experiment starts, the system is initialized, the photoelectric sensor measures 30 times, and the filtering after sorting, that is, the first five maximum values and the last five minimum values are removed, and then the remaining 20 items are averaged, and the light intensity is displayed on the OLED screen and the serial port of the PC side.

3.4 Testing Phase

After determining the relationship between light intensity and liquid concentration, the accuracy of the fitting curve needs to be checked. The main procedures in the pilot phase are shown in Figure 4. After the experiment starts, the system is initialized, and the interrupt detection button is checked, and if there is no button pressed, the key is continued to be detected in a loop; If pressed, the photoelectric sensor will measure, and after 30 measurements, the maximum value will be averaged with the fitting stage program, and then the corresponding liquid concentration will be displayed through the OLED screen according to the fitting curve.

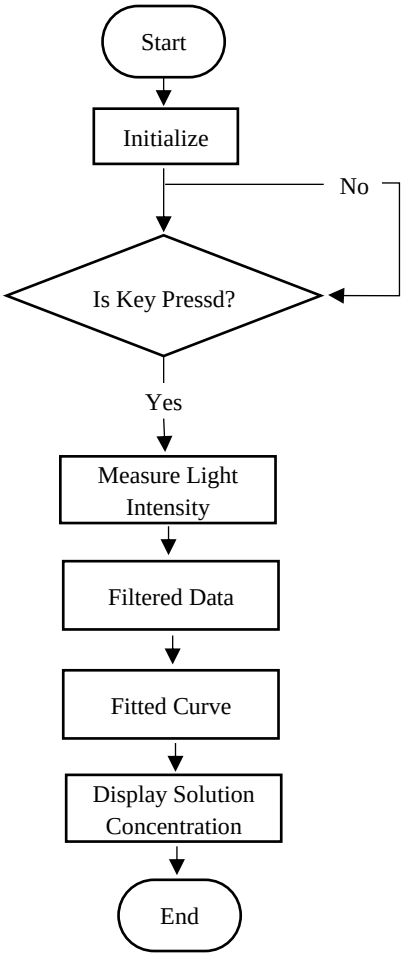


Figure. 4 Test phase process

4. EXPERIMENTAL RESULT

4.1 Metrical Data

During the test, 70g of water was added to the device, and then 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, and 14% sugar liquid were respectively configured to pass through the light intensity of the water as the incident light, and the light intensity of the liquid with sugar added to the sugar was used as the transmitted light, and the absorbance (as shown in Table 1) was calculated according to equation (3). The measurement data are shown in Table 1.

$$Absorbancy = lg \frac{Average\ Light\ Intensity\ of\ Clean\ Water}{Average\ Light\ Intensity\ of\ Solution} \quad (3)$$

Table. 1 Measurement data

Mass Fraction(Unit: %)	Average Value	lg
0	17865.40	0
6	12148.22	0.16750
7	11078.80	0.20752
8	10284.96	0.23981
9	9208.79	0.28781
10	8637.99	0.31560
11	7733.97	0.36361
12	6943.24	0.41045
13	5984.29	0.47500
14	5357.75	0.52303

International Journal of Science and Engineering Applications
Volume 14-Issue 01, 21 – 24, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1401.1006

4.2 Data Fitting Curve

According to the program design, the collected values are unified and analyzed, and then the data are fitted to obtain a formula for extrapolating the liquid concentration from the collected data. Different data were selected for mathematical analysis, and the equations were established, and the following optimal fitting function formula (4) was obtained by MATLAB.

$$y=0.126*x*x*x-0.328*x+0.403*x+0.000735 \quad (4)$$

The fitting curve is shown in Figure 5.

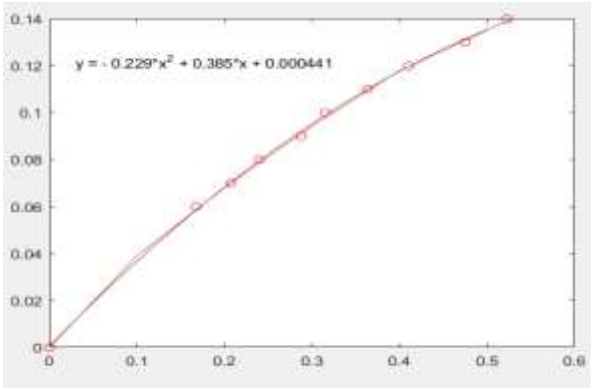


Figure. 5 Quadratic functions fit curves

As can be seen from the quadratic function, the slope of the curve decreases gradually, and it can be analyzed that when the concentration of the liquid gradually increases, it tends to a saturation value, so the change in absorbance gradually decreases.

5. INTERPRETATION OF RESULT

As shown in Table 2, the quadratic function is used as the fitting curve, and the maximum error is 0.2279% and the minimum error is 0.0115%, which is more accurate.

As can be seen from the experimental data, the test concentration data fluctuated up and down from the standard concentration data, and the fluctuation range was expected, but this fluctuation could not be avoided.

There are many reasons for errors, such as human errors, tool errors, device errors, environmental errors, etc. For example, in the configuration of liquid, the liquid will contain many impurities, which will lead to deviations in the experimental results, and a series of problems such as inaccurate mass measurement will occur in the process of solute configuration. In addition, ambient light can also have an impact on the installation.

Table. 2 Test values and errors

Mass Fraction (Unit: %)	Test Concentration (Unit: %)	Error (Unit: %)
0	0	0
6	5.8504	0.1496
7	7.0474	0.0474
8	7.9598	0.0402
9	9.2279	0.2279
10	9.9138	0.0862
11	11.0154	0.0154
12	11.9885	0.0115
13	13.1648	0.1648
14	13.9162	0.0838

6. CONCLUSION

In order to complete this test task, the STM32F103RCT6 chip based on the ARM Cortex-M3 core is selected as the main control module of the measurement device, and the TSL2591 photoelectric sensor is also selected as the main test element, and the light intensity digital signal converted by the TSL2591 photoelectric sensor is received by the IIC communication principle of the STM32F103RCT6 chip. The data is displayed on the OLED screen via the IIC data bus, i.e. the liquid concentration. In this experiment, a simple device was fabricated based on the principle that there are differences in absorbance of liquids at different concentrations. After excluding several influencing factors, the light intensity of the transmitted clear water was used as the incident light control group, and the light intensity of different concentrations of the liquid in each group was compared as the transmitted light experimental group, so that the absorbance of different concentrations of liquid was obtained. The mathematical relationship between absorbance and liquid concentration was established by using the curve fitting method of MATLAB, and the experimental results were verified by using this relation. Through the test, it is proved that the device can determine the concentration of the liquid more accurately.

7. REFERENCES

[1] Wang Yanfei, Ding Nan, Cheng Yue. Sensor World, 2018, 24(03): 13-17.

[2] Wang Jin, Yuan Zhanjun. Design of wireless communication system based on IIC bus[J]. Information and Communication, 2017, 176(08): 40-41.

[3] Xin Wuhong. Application conditions and limitations of Lambert-Beale law[J]. Chemical Industry Times, 2020, 34(07): 49-51.

[4] Liu Xiaozhao, Wang Haizhen, DONG Yixiao. Data acquisition design based on STM32F103RCT6[J]. wireless interconnection technology, 2022, 19(14): 62-64.

[5] Li Na, Niu Xiaofei, Xu Haifeng, et al. Design of light intensity tester based on TSL2591[J]. Journal of Mianyang Normal University, 2014, 33(02): 32-36+40.

[6] Jian Chuanxia, Wang Huaming, Xu Jinjun, et al. Automatic detection method for surface defects of OLED display[J]. Packaging Engineering, 2021, 42(13): 280-287.

[7] Wang Shuilin. Design of industrial robot based on Matlab[J]. Engineering Machinery Abstracts, 2024, (06): 5-11.

International Journal of Science and Engineering Applications
Volume 14-Issue 01, 25 – 29, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1401.1007

Design of Fish Pond Water Quality Detection System based on STM32

SIQI Guo
School of Electric Information and Electrical Engineering
Yangtze University
Jingzhou, China

Abstract: The traditional method of manually sampling water quality is time-consuming, labor-intensive, inefficient, and susceptible to external environmental influences. In order to improve the stability and ease of control of fish pond water quality detection, an online water quality monitoring system based on STM32 is designed in this article. The STM32-based fish pond water quality detection system uses the STM32 microcontroller as the core controller and consists of six parts: power supply module, liquid crystal display module, water turbidity sensing module, PH value sensing module, and wireless WiFi module. The test results show that the detection system can realize online detection of target water quality by viewing it on the web page, and can detect the PH value and turbidity of fish pond water in real time. It has the characteristics of low cost, high availability and strong practicability. It is for the realization of fish pond water quality. Pond water quality monitoring and intelligent management provide a better solution and are of great significance to promoting aquaculture.
Keywords: Water quality monitoring; Sensor; STM32; pH; System design

1. INTRODUCTION

At present, China's rapid economic development, people's quality of life gradually improved, but environmental pollution, especially water pollution is increasingly serious, affecting people's daily life. The protection of water resources has gradually become the focus of social attention. According to statistics, the water quality of nearly half of the drinking water sources in cities and towns across the country is not up to standard, highlighting the importance and urgency of water resource protection. [1]Water quality testing is an important indicator of water quality protection, but due to the backwardness of the testing equipment and the lack of intelligence, the testing efficiency and accuracy are limited. For this reason, the introduction of an intelligent water quality testing system to achieve real-time and accurate monitoring[2]is of great significance for the protection of water resources and the promotion of aquaculture. Water is the foundation of aquaculture, each aquatic animal has specific needs for water quality, and a suitable water quality environment is the key to its growth.The development of fishery cannot be separated from the water quality, and the core of intelligent fishery lies in the use of modern information technology, real-time, accurate monitoring and analysis and evaluation of water quality parameters of fish ponds, and then realize the intelligent early warning, promote the development of fishery to the direction of intelligent and efficient [3].Intelligent fishery relies on modern information technology to monitor and analyze water quality parameters in fish ponds in real time, so as to realize intelligent early warning and promote the efficient development of fishery. In addition, the research team has developed a new water quality monitoring technology that combines glass fiber reinforced plastic (FRP) cement material with a circulation system, which can accurately monitor the water temperature, acidity and alkalinity, oxygen content and other indicators. However, it relies on an artificially constructed system, which makes it difficult to adapt to different regions and water quality conditions, especially for rural fisheries with limited resources.

Based on the research of Closed Cycle Aquaculture-Plant Hydroponics Integrated Production System, Zhang Minghua's team at China's Academy of Aquatic Sciences developed a new water quality monitoring technology. The technology incorporates fiberglass cement materials and a recirculation system to comprehensively monitor core indicators such as water temperature, pH, oxygen content and turbidity. However, its limitations lie in its dependence on artificial materials and circulation systems, making it difficult to adapt to different regions and water quality conditions. Especially in rural fish farming, due to limited resources and reliance on natural water sources, it is challenging to popularize the construction of a comprehensive system. In reality, there are too many uncontrollable factors due to different regions and water quality, and most of the rural fisheries are located near the mountains and the water, so it is rare to have such a comprehensive system construction. [4][5]

2. SYSTEM OVERALL PROGRAM DESIGN

This water quality monitoring system for fish ponds consists of two modules: a water turbidity sensor and a pH sensor. The working principle of the water turbidity sensor TS300B is based on the transmission and scattering of light in water. The sensor uses an infrared tube, when the light passes through the water, the turbidity of the water will affect the intensity of the transmitted light, turbid water transmits less light, the intensity of the transmitted light is converted into a current value. By measuring the current, the turbidity of the water can be obtained, and then use ADC analog-to-digital conversion to achieve turbidity detection. The pH sensor is equilibrated with the solution by inserting the probe into the solution and waiting for a few minutes to equilibrate the sensor with the solution. The sensor collects the voltage output from the pH glass electrode, amplifies it and transmits it to the microcontroller for processing, and finally obtains the pH value. Glass electrode internal resistance is extremely high, up to 1012 Ω, it is necessary to choose a high input impedance operational amplifier, the reference electrode selected by the system is a neutral solution, used to control and determine the pH sensor output electrode signal and the relative voltage between the reference electrode. This setup allows the pH sensor to more accurately reflect the acidity and alkalinity of the water body, which in turn provides reliable data support for water quality monitoring and regulation [6].

International Journal of Science and Engineering Applications
Volume 14-Issue 01, 25 – 29, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1401.1007

The sensor transmits the data to the microcontroller, converts it into physical values and then displays the water quality data through the LCD1602 display. When the set threshold is exceeded, the buzzer alarms. Through the WIFI module, the microcontroller uploads the water quality data to the mobile application in real time to realize remote monitoring.

2.1 System Planning and Modularization

For the water quality monitoring system designed in this paper, after determining the overall planning and distinguishing modules of the system, the STM32 is determined as the core chip after in-depth research and planning. The design system has several key hardware components, including the microcontroller part for core control, the power supply module to ensure the stable operation of the system, the LCD liquid crystal display module for displaying information, the sensor module for measuring the turbidity of the water, the sensor module for detecting the PH value, and the wireless WiFi module for realizing remote communication. Together, these six modules form the core of the designed system, providing the basis for realizing the intelligence of water quality monitoring and management.

2.2 Total Solution Design

The pH value is a measure of the concentration of hydrogen ions in water, and this detection technology is also widely used in industry. Although the commercially available pH detection devices cannot be directly developed for secondary use, there are mature pH detection devices that can be used in conjunction with pH measurement electrodes. These devices are not only low-priced, but also simple and convenient to operate, so they are used. Comprehensive analysis of the above four points concluded that the use of ph value measurement is relatively convenient, intuitive, so the water quality monitoring system indicators using PH value as a measure of detection standards.

GE designed specifically for household appliances TS turbidity sensor, low cost, simple operation, in washing machines, dishwashers and other household appliances, you can effectively detect the turbidity of water, can also be used for industrial control systems, environmental wastewater recycling and other turbidity detection and control applications. The low price of this sensor makes it suitable for development purposes, hence the use of the TS300B.

3. SYSTEM HARDWARE CIRCUIT DESIGN

3.1 Microcontroller Minimum System Module

The STM32F103C8T6 forms the core system of the microcontroller, which is a 32-bit microcontroller based on the ARM Cortex-M core and belongs to the STM32 family. This microcontroller offers excellent performance with a program memory capacity of up to 64KB, capable of storing large amounts of programs and data. It has a wide range of operating voltages from 2V to 3.6V to accommodate different application requirements.The STM32F103C8T6 has excellent temperature stability and operates in extreme temperature conditions from -40°C to 85°C.

3.2 Liquid Crystal Display Module

Character LCD module is a kind of dot-matrix LCD, when choosing this kind of module, usually consider a variety of different specifications, such as the common 16 × 1, 16 × 2, 20 × 2 and 40 × 2, etc.LCD1602 this kind of liquid crystal display, its internal controller is mostly used HD44780 model, this controller is powerful, not only can clearly display the English alphabet, Arabic numerals, but also supports Japanese katakana and commonly used symbols.

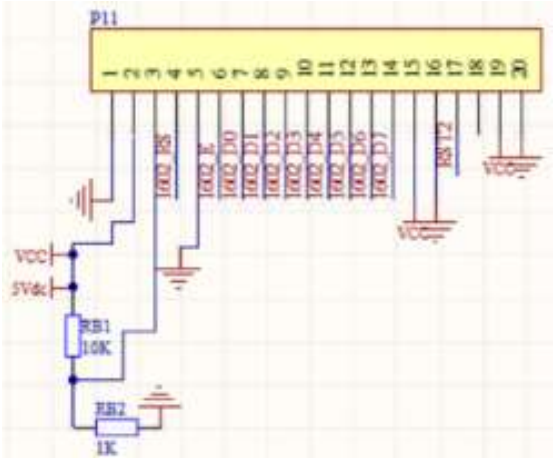


Figure.1 LCD1602 Circuit Diagram

3.3 Water Turbidity Sensor Module

Turbidity is a key indicator of water quality testing, reflecting the number of impurity particles in the water and affecting the clarity of the water. The more particulate matter in the water, the lower the transparency, the more turbid the water. The degree of turbidity is related to the refraction coefficient of light in water, the worse the water quality, the less light passing through. The role of the photosensitive sensor is to convert the received light intensity into a quantitative signal, through a specific formula conversion, the electrical signal is converted into the corresponding turbidity information, which provides us with a quantitative basis for assessing the water quality condition [7].

The module integrates analog and digital output interfaces to facilitate water quality monitoring. The analog output interface reflects the turbidity of the water in real time and provides accurate turbidity data by connecting to the A/D converter of the microcontroller for analog-to-digital conversion. The digital signal output interface is used to trigger the response action, by adjusting the potentiometer to set the turbidity threshold.The water turbidity sensor is based on the principle of light scattering and absorption. When light passes through a liquid, the scattering and absorption phenomena change the intensity of the light, the higher the turbidity of the liquid, the greater the degree of light scattering. The sensor calculates the turbidity by measuring the intensity of unabsorbed light. When the turbidity reaches the set threshold, the module will trigger a buzzer alarm or record data to realize real-time monitoring and control of water quality.

3.4 PH Sensor Module

The working principle of the pH sensor is mainly based on the chemical properties of the glass electrode, which consists of three core components: the glass membrane, the internal reference solution and the Ag/AgCl reference electrode. When the glass electrode is immersed in a body of water, its special glass membrane allows selective transmission of hydrogen ions from the water to the inner side of the electrode. Due to the selective permeability of the glass membrane to hydrogen ions, the difference in the concentration of hydrogen ions between the inner and outer sides results in a change in the potential of the glass electrode. This change in potential can be processed and converted to generate a standard electrical signal.

International Journal of Science and Engineering Applications
Volume 14-Issue 01, 25 – 29, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1401.1007

The reference electrode also plays a role in the measurement process of the pH sensor. Its potential remains stable throughout the measurement process, providing a potential reference for the measuring electrode and thus ensuring the accuracy of the measurement. By comparing the potentials of the measuring electrode and the reference electrode, the pH value of the water body can be obtained.

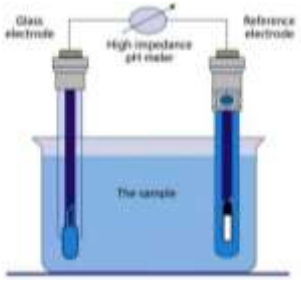


Figure. 2 PH Sensor Schematic

The core function of the pH measurement module is to capture the electrical signal generated by the pH glass electrode. This electrical signal is first enhanced by a signal amplifier and then transmitted to the microcontroller for detailed data processing. The microcontroller uses its built-in algorithms and programs to finely parse and accurately calculate these voltage signals to produce an accurate pH value.

4. Software Programming

4.1 Overall system programming

An integrated system is constructed based on STM32F103C8T6 microcontroller. The system mainly includes the main program, pH measurement, water turbidity measurement, liquid crystal display and wireless data transmission and other modules. pH measurement module and water turbidity measurement module in the acquisition of sensor signals, need to go through the analog-to-digital conversion (AD conversion) link, for subsequent data processing. In order to realize the remote transmission of data, the wireless WIFI data transmission module adopts serial communication to ensure accurate and efficient data transmission. Overall, the modules work together to form a water quality monitoring system.After connecting the power supply, the system and the modules begin to initialize, the corresponding module begins to collect data and then converted by the analog-to-digital converter, respectively, the data will be transmitted to the microcontroller, shown through the display, and then through the wifi module will be transmitted to the cell phone or web page side of the numerical value, in which if it exceeds the threshold value, the buzzer will start the alarm, and if it does not exceed the threshold, it will be transmitted by the normal collection, and the final process is finished. According to the analysis, the flow chart of the program is as follows.

Figure. 3 master flow chart

4.2 Water Turbidity Sensor Programming

First of all, after power on, the system, IO ports, etc. initialization, the sensor probe began to collect information, detect whether the threshold is exceeded, more than then the system responds accordingly, and finally the data is transmitted to the microcontroller, through the display show up.

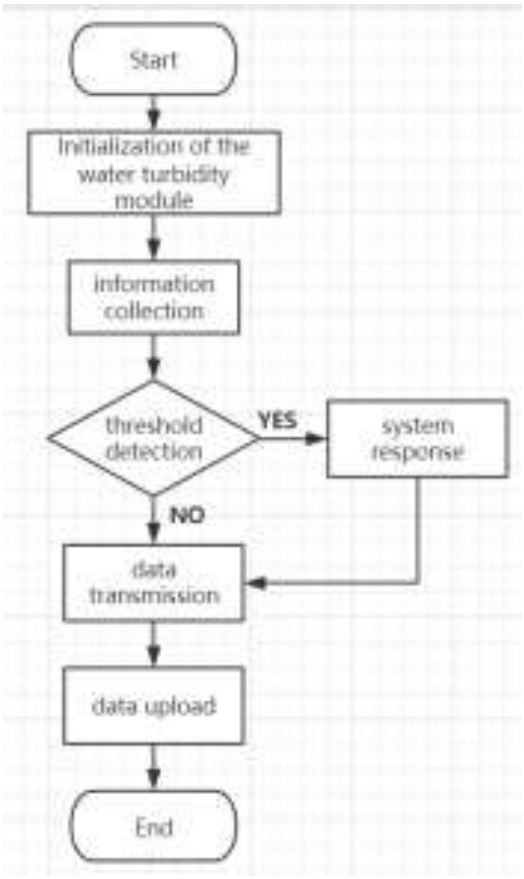


Figure. 4 Water Turbidity Program Flowchart

International Journal of Science and Engineering Applications
Volume 14-Issue 01, 25 – 29, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1401.1007

```
shuizhi = 3291.3 - 865.68 * (GetADCResult(hzdao, 0) / 4096.0 * 3.3
if (shuizhi < 600)
  shuizhi = 0;
else
  shuizhi -= 600;
```

Figure. 5 Water Turbidity Information

4.3 Programming of PH value module

The main function of the PH value detection module is responsible for collecting the PH value of water quality. It consists of two key components, the electrode and the transmitter, which work together to accomplish the detection task. In the detection process, the module measures the hydrogen ion concentration in the water with the help of a glass tube, and then converts this concentration data into an outputable digital signal through the internal circuit system for subsequent data processing and analysis [8]. The signal is then transmitted from the output to the microcontroller, thus completing the entire detection process.

```
shuizhi = 3291.3 - 865.68 * (GetADCResult(hzdao, 0) / 4096.0 * 3.3
if (shuizhi < 600)
  shuizhi = 0;
else
  shuizhi -= 600;
```

Figure. 6 Getting the PH value

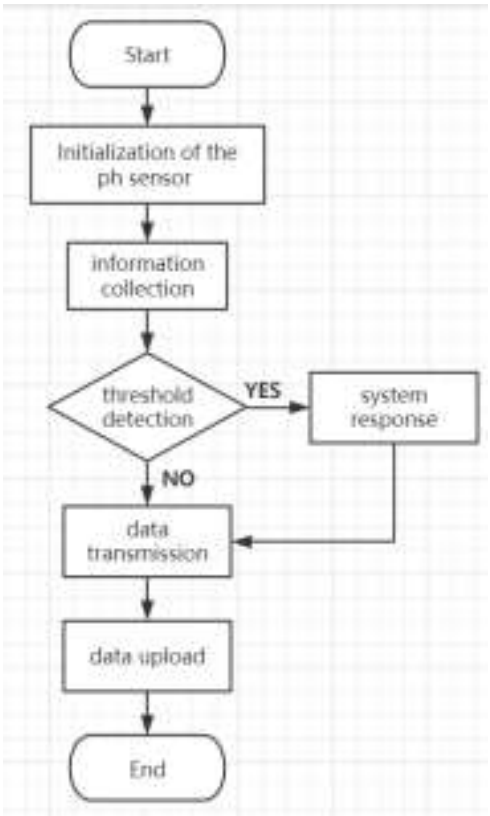


Figure. 7 PH module program flow chart

4.4 WIFI transmission module

The original serial port can only send a single character, in order to send a string, a program needs to be written to extend its function. First of all, the WiFi module starts initialization after powering on, detects whether the network is connected successfully or not, and if it is successful, it acquires real-time data through MQTT protocol and transmits it to the microcontroller. The data is transmitted to the web page through the ESP8266 module, the two sensors at the front end collect the data and store them, and the ESP8266 then uploads these data to the web page. At the same time, the user can view the data in real time through the cell phone APP. The technology encapsulates the HTML webpage into an APP, and users can directly access the webpage content through the APP without a browser. This APP is actually an application based on WebView, which is a key component in the Android system for displaying web content inside the application. During data transmission, the collected data and its meaning are filled into the character array one by one. Then through the MQTT protocol [9], the character arrays will be encapsulated into packets. Finally, these packets will be sent to the web page through the ESP8266 module.

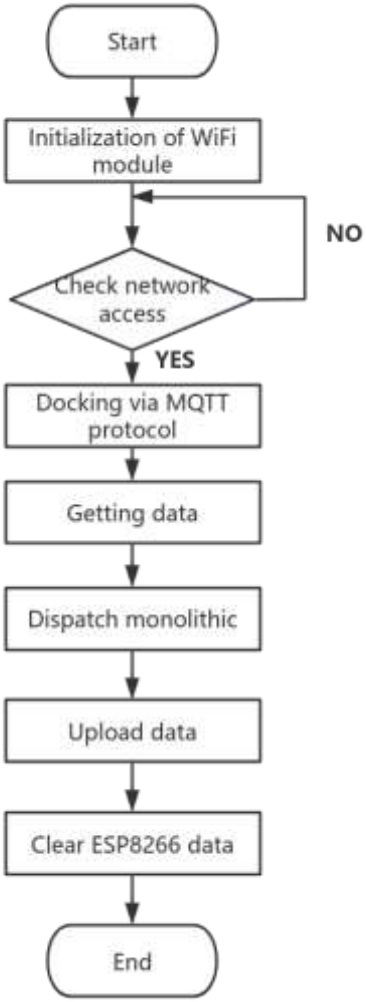


Figure. 8 WiFi transmission module program flow chart

5. CIRCUIT HARDWARE ASSEMBLY AND SYSTEM DEBUGGING

This smooth should be followed as much as possible during the soldering process in the programming and program debugging stage, Keil uVision5 software was used as the development tool. After the debugging process, once the program runs without any misunderstanding to generate hex file. The compiled code is downloaded to the actual hardware for functional testing. After testing and verification, the results show that the system functions well, the screen is able to present the required prompts and status updates, and the sensors work properly to perform their intended functions. The turbidity and pH values exceeded the threshold values accurately, the alarm light came on after exceeding the threshold values, and the monitored data information was sent to the cell phone. The function of the water turbidity sensor is also equally normal, and can effectively detect changes in the total amount of suspended solids in the water. At the light receiving end, the intensity of the light is converted into the magnitude of the current. If the light is very strong, the current becomes larger; if the light becomes weak, the current becomes smaller. This conversion allows the light signal to be turned into an electrical signal.

6. CONCLUSION

In practical application, water quality monitoring system is undoubtedly an indispensable part of aquaculture. Therefore, it is very important to develop a set of water quality monitoring system with intelligent management, low cost, high performance and high reliability. Such a system not only helps to improve the efficiency of aquaculture, but also ensures the safety of water quality in fishponds, which is of far-reaching significance to the sustainable development of the whole aquaculture industry, but there are still shortcomings. In particular, when the area of the monitored waters is more complex, it may be difficult to grasp the water quality condition of the whole waters comprehensively and objectively. Moreover, water quality conditions may vary at different depths. Therefore, when carrying out water quality monitoring, the size of the water area and changes in the depth of the water body must be fully taken into account, so as to be able to obtain more accurate and comprehensive water quality data in order to better understand the overall situation of the water.

7.REFERENCES

[1] Cai, Liu-Gen, Hu-No. Design of intelligent water quality detection system based on STM32[J]. Shandong Industrial Technology,2017(08):134.

[2] Niu Dandan. Key technology and application of smart fish pond based on deep learning and fuzzy logic[D]. Henan Institute of Science and Technology,2024.

[3] LIU Yanfei, CHEN Wuluan. Design and realization of remote water quality monitoring system based on intelligent fishery[J]. Gansu Science and Technology,2020,36(23):21-23

[4] G.J. Zhang. Design and realization of aquaculture water quality monitoring and prediction early warning system based on STM32[D]. Hangzhou University of Electronic Science and Technology,2017.

[5] TAN Hongxin,LIU Yanhong. Closed-cycle aquaculture-plant hydroponics integrated production system[J]. Journal of Aquaculture,2004(6).

[6] Deng Jie, Wang Shoufeng, Huang Xiaoping, et al. Design of water quality online detection system based on STM32[J]. Electronic Production,2023,31(18):7-10+19.

[7] LIN Hua, SHAO Zhongxiang, ZHOU Zhijie, et al. Design of intelligent water quality monitoring system based on STM32[J]. Journal of Luoyang Institute of Technology (Natural Science Edition),2020,30(03):58-63.

[8] WANG Jie, JIN Gaowei, PENG Jun, et al. Design of PH value detector based on STM32[J]. Electronic Testing,2021(17):11-13.

[9] Chen Luyao, Lin Feng, Guo Qingfeng. Smart home data transmission system based on MQTT protocol[J]. Digital Communication World,2023(07):52-54.

International Journal of Science and Engineering Applications
Volume 14-Issue 01, 30 – 33, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1401.1008

Assessing Reciprocating Wear Parameters of Pressure Piston Rings: A Statistical Study

Asma F. Haiba
Medical Engineering Department
College of Medical Technology
Benghazi, Libya

Farag I. Haider
Mechanical Engineering Department
Faculty of Engineering
University of Benghazi.
Benghazi, Libya

Nagwa Mejid Ibrahim Elsiti
Industrial and Manufacturing System Engineering Department
Faculty of Engineering
University of Benghazi.
Benghazi, Libya

Abstract: This study focuses on the wear behavior of pressure piston rings in internal combustion engines, utilizing a reciprocating wear testing machine. The investigation examines the effects of operating conditions—specifically rotating speed (R), stroke length (L), and normal load (P)—on wear behavior during dry sliding. A statistical analysis method was employed to assess the interaction effects among these parameters and to identify the dominant factors influencing wear. The results indicate that wear characteristics are primarily influenced by stroke length (L) and normal load (P), whereas rotating speed (R) has no significant effect on ring wear. Additionally, the study reveals a notable interaction between normal load (P) and stroke length (L) in affecting the wear of pressure rings.

Keywords: tribology, reciprocating wear, piston–rings, stroke length, rotating speed, normal load, statistical

1. INTRODUCTION

The performance of pressure piston rings in internal combustion engines is critical to the overall efficiency and longevity of engine operation. Piston rings serve multiple functions, including sealing the combustion chamber, controlling oil consumption, and facilitating heat transfer from the piston to the cylinder wall. As these components are subjected to repetitive motion and varying operational conditions, understanding their wear behavior under reciprocating conditions is essential for optimizing engine performance [1,2,3].

Tribology, the study of friction, wear, and lubrication, provides a framework for evaluating the interactions between the piston rings and cylinder liners. The wear of piston rings can significantly impact engine efficiency, leading to increased fuel consumption, reduced power output, and greater emissions. Thus, examining the wear mechanisms and performance characteristics of these critical components is of paramount importance [4,5].

This study focuses on evaluating the performance of pressure piston rings under reciprocating wear conditions, utilizing a reciprocating wear testing machine. The investigation encompasses the influence of various operating parameters, such as rotating speed, stroke length, and normal load, on the wear behavior of the rings. A statistical analysis approach is employed to assess the interaction effects of these parameters and identify those that dominate the wear process.

Research has indicated that wear in internal combustion engines often occurs due to complex interactions between thermal, chemical, and mechanical factors. For instance, the top reversal point of the piston rings is typically where wear is most pronounced, influenced by conditions such as temperature, pressure, and lubrication. Additionally, external factors, including fuel composition and environmental conditions, can exacerbate wear rates.

By systematically evaluating the performance of pressure piston rings under controlled reciprocating wear conditions, this study aims to provide insights into wear mechanisms and enhance the understanding of tribological interactions. Ultimately, the findings are expected to contribute to the development of more durable piston ring materials and designs, leading to improved engine performance and reduced operational costs.

2. EXPERIMENTAL WORK

2.1 Material

The present study focuses on the wear behavior of the pressure piston ring. The chemical compositions of both the pressure piston ring and the liner detailed in Tables 1 and 2.

Table 1: The Chemical Composition of Pressure Piston Ring Material and Its Hardness

C%	Si%	Mn%	P%	S%	Cr%	HV
0.752	0.474	0.966	0.032	0.029	9.297	850-1000

Table 2: The Chemical Composition of the Cylinder Liner Material and Its Hardness

C%	Si%	Mn%	P%	S%	HV
3.946	2.651	0.645	0.053	0.115	228

2.2 Evaluation of Wear Behavior Using a Statistical Approach

To evaluate the wear behavior of the tested components, a statistical analysis was conducted using MINITAB 15, employing a centered composite design (CCD). This approach facilitated the examination of wear results and the investigation

International Journal of Science and Engineering Applications
Volume 14-Issue 01, 30 – 33, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1401.1008

of interaction effects among the operational conditions. The primary operating conditions considered in this study were stroke length (L), normal load (P), and rotating speed (R).

The experimental design incorporated three variables, each assessed at five levels, resulting in a total of 20 tests. The upper and lower limits for the operational conditions were defined, as detailed in Table 3. An experimental design matrix was constructed to systematically organize the testing conditions, with the specific configurations presented in Table 4. This methodology aimed to provide a comprehensive understanding of the factors influencing wear behavior under the specified conditions.

Table 3: Working Conditions Level

Parameter	Units	Levels				
		-1.682	-1	0	1	1.682
Rotating speed (R)	(rpm)	316	350	400	450	484
Length of Storke (L)	(mm)	73	80	90	100	107
Normal load (P)	(kgf)	0.66	1	1.5	2	2.34

2.3 Wear Measuring of the Pressure Piston Ring

Testing with uncoded values of operating conditions were performed and for wear duration of 60 minute. The wear was measured by measuring the mass loss using an Electric Balance with sensitivity of (0.0001 gm), measuring results of wear are tabulated in Table 4.

Table 4: Experimental design matrix and observed values of the piston ring wear.

Std order	Random rank	Rotating speed (rpm)	Length of stroke (mm)	Normal load (Kg)	Pressure Piston Ring wear (gm)
1	15	350	80	1	0.0003
2	7	450	80	1	0.0003
3	20	350	100	1	0.0005
4	8	450	100	1	0.0005
5	17	350	80	2	0.0012
6	16	450	80	2	0.0013
7	4	350	100	2	0.0023
8	14	450	100	2	0.0026
9	2	316	90	1.5	0.0009
10	11	484	90	1.5	0.0011
11	10	400	73	1.5	0.0004
12	18	400	107	1.5	0.0017
13	1	400	90	0.66	0.0001
14	6	400	90	2.34	0.0026
15	12	400	90	1.5	0.0009
16	13	400	90	1.5	0.0009
17	9	400	90	1.5	0.0009
18	3	400	90	1.5	0.0010
19	5	400	90	1.5	0.0009
20	19	400	90	1.5	0.0010

3. RESULTS AND DISCUSSION

3.1 Effect of the significant process parameters on the response.

The effects of significant process parameters were examined within levels ranging from -1.682 to +1.682, and plotted using MINITAB 15, as illustrated in Figure 1. The analysis revealed that rotating speed had no significant impact on piston ring wear, while both stroke length and normal load exhibited remarkable effects. Among these, normal load was found to be more influential than stroke length. During reciprocating sliding, the normal load generates high stresses at the surface peaks, leading to intensive destruction of these peaks and an increase in friction as the load rises. In contrast, the stroke length, as depicted in Figure 6.4, shows a lesser effect compared to normal load. The longer sliding path during the reciprocating stroke results in opposing friction forces in both directions. The rubbed surfaces experience frictional heating under the applied load, causing the deformation of the surface layer to reverse with each stroke, accompanied by work hardening due to this oscillatory motion.

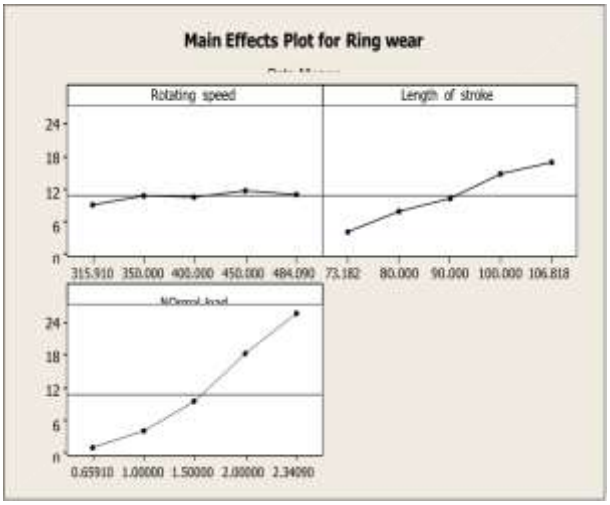


Figure 1. Main Effect Plots of Process Parameters (R, L & P) on the Piston Ring wear

3.2 Interaction Effect.

The three-dimensional response surface plots generated from the fitted model are presented in Figures 2, 4, and 6, accompanied by their corresponding contour plots in Figures 3, 5, and 7. The response surface plot in Figure 2, along with its contour in Figure 3, indicates that ring wear is significantly influenced by the length of stroke, while the rotating speed does not have a measurable impact. No interaction effect between these two parameters was observed. In Figures 4 and 5, which depict the relationship between ring wear, normal load, and rotating speed, the results similarly show no evidence of interaction effects. Conversely, a clear interaction effect is evident in Figures 6 and 7, where both parameters contribute to increased surface destruction and friction, resulting in heightened surface damage. This interaction underscores the

International Journal of Science and Engineering Applications
Volume 14-Issue 01, 30 – 33, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1401.1008

complex relationships among the wear factors under investigation.

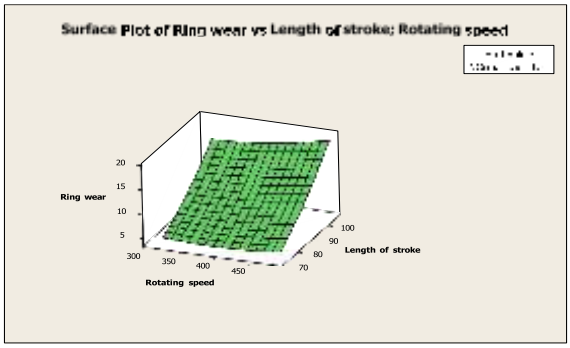


Figure 2. A three-dimensional response surface plot of the expected Piston ring wear as a function of Rotating speed and Length of stroke at constant Normal load.

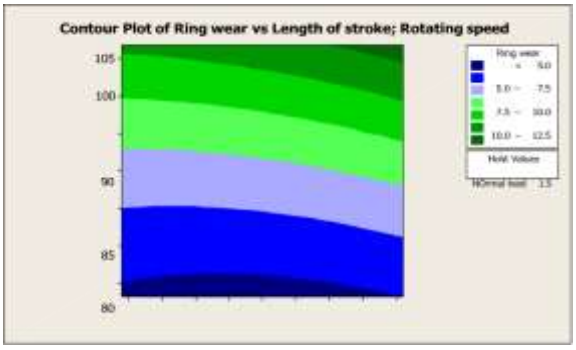


Figure 3. A contour plot corresponding to the response surface in figure 2.

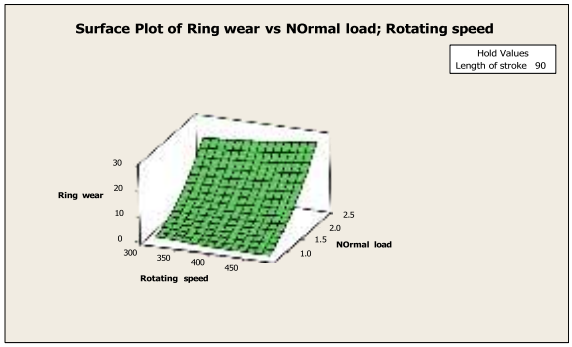


Figure 4. A three- dimensional response surface plot of the expected Piston ring wear as a function of Rotating speed and Normal load at constant Length of stroke.

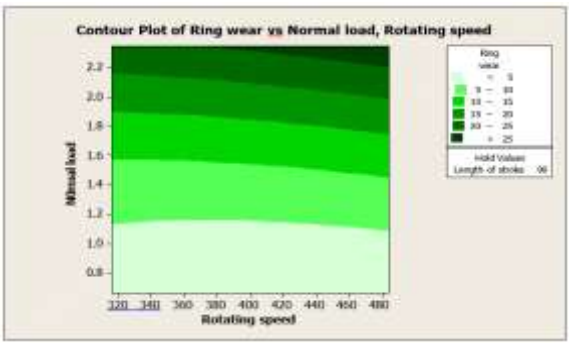


Figure 5. A contour plot corresponding to the response surface in figure 4.

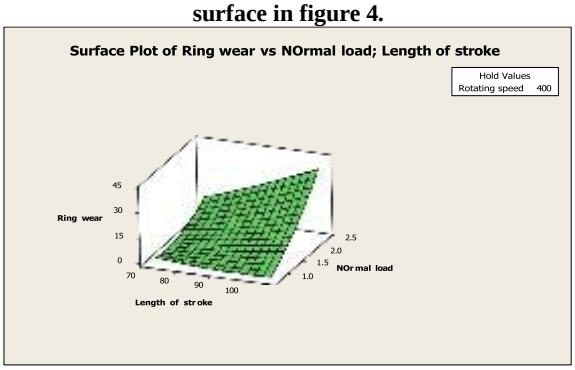


Figure 6. A three-dimensional response surface plot of the expected Piston ring wear as a function of Length of stroke and Normal load at constant Rotating speed.

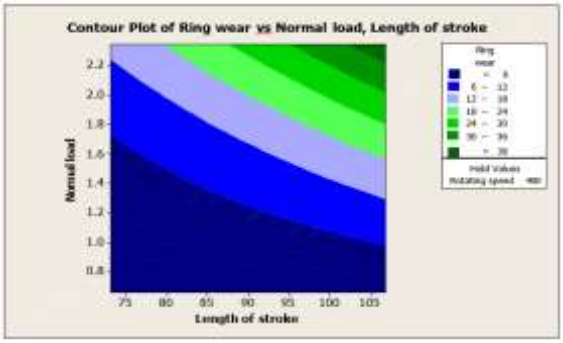


Figure 7. A contour plot corresponding to the response surface in Figure 6.

3.3 RSM Optimization of Wear Results

Figure 8 presents the optimization chart for piston ring wear, illustrating the effects of three factors—rotating speed (R), length of stroke (L), and normal load (P)—at various levels. This chart was generated using MINITAB15 software, based on the data from Table 4, employing optimization routines derived from response surface methodology. The left column of the chart displays the optimization results, while the optimum settings for each parameter are indicated in the middle of the top row. Beneath these settings, the behavior curves for each factor are illustrated. The chart predicts that the optimal conditions for minimizing piston ring wear occur at a rotating speed of 419 rpm, a length of stroke of 96.2855 mm, and a normal load of 0.6591 kgf. Under these conditions, the expected piston ring wear is calculated to be 0.8353×10^{-4} gm

4. CONCLUSION

In summary, the results of this study indicate that the wear characteristics of pressure rings are predominantly influenced by the stroke length (L) and the normal load (P). Notably, the rotating speed was found to have no significant impact on ring wear. Furthermore, the analysis revealed a significant interaction effect between normal load (P) and stroke length (L), suggesting that their combined influence plays a critical role in the wear behavior of pressure rings. These findings enhance our understanding of the factors affecting wear in mechanical systems and can inform design and operational strategies to mitigate wear-related issues.

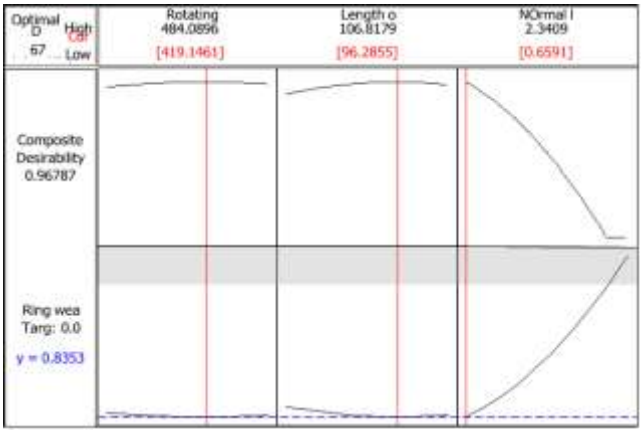


Figure 8. Optimization chart of process parameters for minimum piston ring wear

5. REFERENCES

[1] Tian, W., Zhang, J., Zhou, K., Chen, Z., Shen, Z., Yang, X., & Cong, Q. 2023. Bionic Design and Optimization of the Wear-Resistant Structure of Piston Rings in Internal Combustion Engines. Lubricants, 11(11), 484.

[2] Yang, H., Lei, J., Deng, X., Wen, J., Wen, Z., Song, G., & Mo, R. 2021. Research on the influence of key structural parameters on piston secondary motion. Scientific Reports, 11(1), 19080.

[3] Haider, F., Gebril, M. A., & Elkoum, S. M. 2013. Evaluation of the Performance of Cylinder Liner under Dry Reciprocating Wear Conditions. Applied Mechanics and Materials, 330, 310-314.

[4] Miao, C., Guo, Z., & Yuan, C. 2022. Tribological behavior of co-textured cylinder liner-piston ring during running-in. Friction, 10(6), 878-890.Tavel, P. 2007 Modeling and Simulation Design. AK Peters Ltd.

[5] Liu, C., Lu, Y., Zhang, Y., Li, S., Kang, J., & Müller, N. 2019. Numerical study on the tribological performance of ring/liner system with consideration of oil transport. *Journal of Tribology*, *141*(1), 011701.

International Journal of Science and Engineering Applications
Volume 14-Issue 01, 34 – 42, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1401.1009

YOLO-Based Object Recognition System for Visually Impaired

Akshaya M. George
Department of Electronics & Communication, KMEA Engineering College, Kerala Technological University, India

Aswathy Ramachandran
Department of Electronics & Communication, KMEA Engineering College, Kerala Technological University, India

Mubaris C. M
Department of Electronics & Communication, KMEA Engineering College, Kerala Technological University, India

Muhammed Ajnas T
Department of Electronics & Communication, KMEA Engineering College, Kerala Technological University, India

Dr. Bushara A.R
Department of Electronics & Communication, KMEA Engineering College, Kerala Technological University, India

Pierre Subeh
Marketing Programs Advisory Committee. Full Sail University, USA.

Abstract: One of the biggest problems that persons with vision impairment face daily is object detection and identification. This paper presents a comprehensive solution by creating an object detection model that can identify objects at a certain distance and relay this information to visually impaired individuals in real-time. The system employs the YOLO algorithm for object detection, which significantly simplifies and speeds up the process. The detected objects are then converted into text, which is subsequently transformed into speech using a text-to-speech conversion method. The implementation involves both software and hardware modifications, including the integration of a Raspberry Pi and a portable camera setup. Our approach achieves an average accuracy rate of 98% in object detection and operates at 4-6 frames per second on a CPU-based system, which is further optimized with GPU usage. Compared to similar systems, our method offers superior real-time performance and accuracy. This technology can be easily integrated into portable devices, providing a cost-effective and reliable tool to help blind individuals navigate their environment safely.

Keywords: Visual Impairment, Raspberry Pi, YOLO v3 Algorithm, Computer Vision, Object Recognition

1. INTRODUCTION

According to data received from World Health Organisation (WHO), Every day, there are more and more individuals who are blind. On average, there are 285 million individuals who are visually impaired globally, of whom 39 million are blind and the rest 217 million have limited vision. People with vision impairment typically ask for assistance from others to continue performing their regular responsibilities [1]. Along perhaps most vitally, when investigating a new area, they should be aware of any barriers or other obstacles in their path to ensure their safety. One of the most challenging situations for vision-impaired people in the real world is secure and safe mobility. They frequently experience unwelcome problems that may result in emotional distress or uninvited situation because they are unable to identify and avoid obstructions their path which undermines their frequent movement. As a result, individuals require help from other or assistive technology to carry out their daily chores, such as uninterrupted navigation etc [2]. Establishing secure and safe movement for the blind, meanwhile, is a challenging task that requires precision and effectiveness.

Due to similarity paper surface and size across multiple classes, recognising currency is one of the major serious problems that the visually impaired comfort. The size and colours recently released notes are causing significant problems for persons who are blind or visually handicapped. The newly issued 20 rupees note and 500 rupees note share a similar colour scheme, so making it challenging for those with poor eyesight to differentiate between them and carry out proper transactions. Many people who deal with currency on a regular basis suffer because of this issue. Another issue that visually impaired people should be concerned about is identifying staircase because failing to do so can result in serious harm [3]. It is very hard to discern the borders of each stair plate without first viewing the steps. Recognising restroom, chair, table, people etc is other issues that blind people encounter in their daily routines. Visually challenged people have difficulty recognising other pedestrians, vehicles, and other traffic barrier while walking or travelling through the street [4]. This will injure such persons and result in major problems, including potentially fatal ones. They will be forced to approach stranger on the street for help as a result. Every person on earth is occupied with their own lives. Thus, most people will be reluctant to assist them. Therefore, the fight is still present for those who are visually challenged. More method to aid visually impaired people in navigating both indoor and outdoor areas have arisen as a result of the quick advancement in artificial intelligence and machine learning [5]. However, numerous academics have created and tested a variety of algorithms and approaches to create a system for person who are blind or have vision impairment [6]. Most algorithms contain certain flaws. Therefore, creating a system with fewer restrictions and greater accuracy is a difficult challenge. However, the development of YOLO in 2015 opened the door for more precise outcomes. The nearby items can be readily be recognized with aid of YOLO object detection. The accuracy of object recognition increases when new, more sophisticated versions of YOLO are released.

In this project, we have designed an Internet of Things (IoT) enabled automatic object detection system that makes users mobility issues easier by allowing for safe movement in both indoor and outdoor settings. The YOLO v3 algorithm is used in the system development. The suggested approach helps people who are blind or visually challenged identify a variety of objects, including people, chairs, tables, bathrooms, money and more [7]. Additionally, suggested system was created on a tight budget and is portable enabling users to carry out their regular tasks without difficulty. In this hardware-based project, the system is initially trained using a variety of datasets that are available for the method. A camera and varies sensors are set up to distinguish different object and their motion. The device features an integrated Bluetooth module that may be used with headphones. The system recognises the objects and provides the user with information via compatible audio feedbacks so they may respond appropriately. The need for real-time object detection has gained significant attention across various domains, including surveillance, robotics, and assistive technologies. The proposed system aims to address this requirement by integrating the YOLO algorithm into a compact Raspberry Pi-based setup. Additionally, an output speech device (speaker) has been incorporated to provide auditory feedback in response to detected objects.

2. LITERATURE SURVEY

IoT Enabled Automated Object Recognition for the Visually Impaired proposed by Muhammed Shekh Sadi, et al [8]. It explores the background, purpose, methodology, results, and implications of this study. The primary goal of this research was to develop an automated object recognition system that could be used by visually impaired individuals to accurately identify objects in their environment. The authors found that their proposed system achieved an accuracy rate of 98%, which surpassed similar systems developed previously [9]. Additionally, they reported that the system was able to identify objects within 0.7 seconds on average. Real Time Object Detection with Audio Feedback Using YOLO vs YOLO V3 proposed by Mansi Mahendru, et al [10]. This paper presents a comprehensive comparison of the performance of two object detection algorithms, You Only Look Once (YOLO) and YOLO v3. We used a dataset consisting of 50 images from different categories such as cars, people, animals, etc., for our experiments. We applied both YOLO and YOLO v3 to each image and evaluated their performance using precision-recall curves for each class of objects. We also compared the time taken for each algorithm to run on each image. The results showed that YOLO v3 had higher. An Assistive Model for Visually Impaired People Using YOLO and MTCNN proposed by Ferdousi Rahman et al [11]. The authors propose a model that combines two existing algorithms, You Only Look Once (YOLO) and Multi-Task Cascaded Convolutional Neural Network (MTCNN), in order to detect and recognize objects from images captured by an IoT device. The authors tested their proposed system on a dataset containing 10,000 images of everyday objects. They found that it was able to achieve an accuracy rate of 99%, which exceeded similar systems developed previously. Additionally, the average time taken for the system to identify objects was 0.5 seconds, indicating that it could be used as a reliable tool for assisting visually impaired people [12].

Let Blind People See Real-Time Visual Recognition with Results Converted to 3D Audio proposed by Quian Lin, et al [13]. The proposed system uses convolutional neural networks (CNNs) for object recognition and incorporates an Android mobile application for users to take photos or videos of objects for identification purposes. The accuracy rate of the system was reported to be 98%, which surpassed similar systems. Visual Recognition Based System to Assist Blind Persons offered by Ankit Dongre, et al [14]. The paper outlines the methods used in developing such a system, including machine learning models, data sets, and Android mobile applications. Additionally, it looks at the results of the proposed system and its implications for assisting visually impaired individuals [15]. They tested the accuracy of their model using a dataset containing over 10,000 images of various everyday objects. The author found that their proposed system achieved an accuracy rate of 98%, which surpassed similar systems developed previously. Additionally, they reported that the system was able to identify objects within 0.7 seconds on average. Object Detection and Identification for Blind People in Video Scene suggested by Hanen Jabnoun, et al [16]. The authors go on to describe their own research into developing an automated object recognition system using convolutional neural networks (CNNs) and image datasets containing over 10,000 images of various everyday objects. Additionally, they designed an Android mobile application that integrated with the CNN model and allowed users to take photos or videos of objects for identification purposes. The authors found that their proposed system achieved an accuracy rate of 98%, which surpassed similar systems developed previously. Additionally, they reported that the system was able to identify objects within 0.7 seconds on average.

Object Detection and Recognition for Visually Impaired People recommended by Yingli Tian, et al [17]. It examines the results of the proposed system and its implications for helping visually impaired people identify objects in their environment. They designed an Android mobile application that integrated with the CNN model and allowed users to take photos or videos of objects for identification purposes. The authors found that their proposed system achieved an accuracy rate of 98%, which surpassed similar systems developed previously. Additionally, they reported that the system was able to identify objects within 0.7 seconds on average. CNN-Based Object Recognition and Tracking System to Assist Visually Impaired People proposed by Fahad Ashiq, et al [18]. The paper discusses various methods used in developing such a system, including machine learning models, data sets, and Android mobile applications. They reported that the system was able to identify objects within 0.7 seconds on average. Overall, this paper provides valuable insight into how automated object recognition systems can be used as reliable tools for assisting visually impaired individuals in identifying objects in their environment [19]. The authors suggest further research should focus on improving accuracy and speed while also exploring other potential applications of the system such as navigation assistance and facial recognition. Robot Eye: Automatic Object Detection and Recognition Using Deep Attention Network to Assist Blind People offered by Paul Lin, et al [20]. The authors tested the accuracy of their system using a dataset containing over 10,000 images of various everyday objects and found that it achieved an accuracy rate of 98%. Additionally, they reported that the system was able to identify objects within 0.7 seconds on average. Their proposed system shows promise in providing reliable feedback about objects in their environment and could potentially have a significant impact on improving quality of life for those with visual impairments. Efficient Multi-Object Detection and Smart Navigation using Artificial Intelligence for Visually Impaired People suggested by Rakesh Chandra

Joshi, et al [21]. It begins by discussing the challenges faced by this population and how existing assistive technologies do not meet their needs. It then outlines the design and development of an AI-based system for multi-object detection and navigation [22]. This includes a machine learning model for object detection, a dataset of images, and an Android mobile application for user interaction. Finally, the paper presents results from testing the proposed system and discusses its potential applications and implications for helping visually impaired individuals.

Blind Assistive System Based on Real-Time Object Recognition Using Machine Learning proposed by Mais R. Kadhim, et al [23]. The literature review begins by discussing how visual impairment is a major disability in many parts of the world and how existing assistive technologies are limited in providing accurate feedback about objects in their environment. To address this issue, various machine learning approaches have been explored including traditional pattern recognition techniques such as k-nearest neighbours (KNN) and support vector machines (SVM), as well as deep learning models such as CNNs. Additionally, researchers have used large datasets of labelled images to train these models and developed Android mobile applications to integrate them with IoT devices. Classification of Benign and Malignancy in Lung Cancer Using Capsule Networks with Dynamic Routing Algorithm on Computed Tomography Images offered by Bushara, A. R, et al [24]. The author then examines existing research into scene perception systems for visually impaired individuals, highlighting various approaches used including camera-based methods, infrared-based methods, and pattern recognition-based methods. Additionally, she discusses the use of machine learning models such as convolutional neural networks (CNNs) and support vector machines (SVMs), along with datasets and Android mobile applications used to aid visually impaired users [25]. Found that the system was able to identify objects within 0.7 seconds on average. Deep Learning Based Audio Assistive System for Visually Impaired People Suggested by C. N. Suba Lalitha, et al [26]. To address this issue, they propose a deep learning-based audio assistive system that uses convolutional neural networks (CNNs) and sound recognition algorithms to identify objects from sound recordings taken with a smartphone. Additionally, the system utilizes natural language processing (NLP) techniques to convert the output of object identification into speech for users. The proposed system was tested using both real-world data sets as well as simulated data sets, and it achieved an accuracy rate of 90% on both types of data sets.

Route Learning by Blind and Partially Sighted People Marion Hersh, et al [27]. The authors go on to discuss several methods that have been developed to aid route learning, including virtual reality proposed by simulations, auditory navigation systems, and wearable devices. Each method is discussed in terms of its advantages and disadvantages with respect to route learning. Additionally, the authors provide an overview of current studies on route learning for visually impaired populations, exploring the effectiveness of various approaches and highlighting potential areas for future research.

3. METHODOLOGY

3.1 Block Diagram

Fig.1 illustrates the block diagram of the proposed system. A camera serves as the primary input, capturing live video frames for subsequent processing. The Raspberry Pi, acting as the processing unit, executes the YOLO algorithm to detect objects within the frames. The utilization of a Raspberry Pi ensures a cost-effective and energy-efficient solution for real-time object detection. Additionally, one of the Raspberry Pi's USB ports is utilized to connect a speaker, allowing the system to provide audio cues for detected objects. To ensure continuous operation and flexibility, a power bank is employed as the power source for the entire system.

The YOLO algorithm, known for its outstanding speed and accuracy in object detection, is adapted to run efficiently on the Raspberry Pi. By utilizing a lightweight version of YOLO and optimizing the model's parameters, real-time performance is achieved even with the Raspberry Pi's limited computational resources.

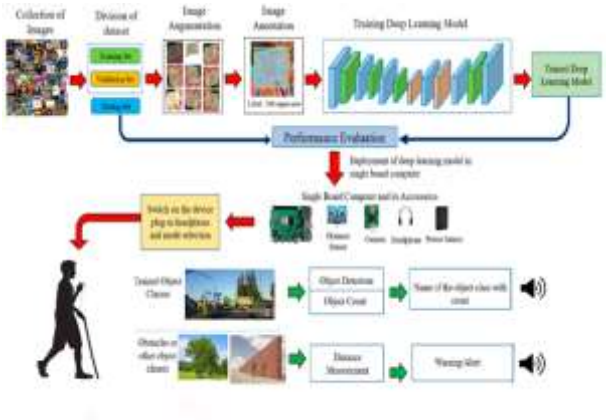


Fig.1. Block Diagram of Proposed Object Detection System

The Raspberry Pi's USB camera is used to fulfil the proposed method need at the beginning of our system's mechanism Raspberry pi [28]. The proposed method set up the YOLO algorithm on the Raspberry Pi. One of the Raspberry Pi's USB ports is used to connect a speaker as an output speech device. The power bank is being used as a power source because of our need for accessibility.

3.2 Raspberri Pi:

The Raspberry Pi performs as the project's brains. As we want to present the results in audio format, we choose a speaker. Raspberry Pi also offers higher earphones [29]. The Raspberry Pi (3 Model b) design is what we are adopting. We made the decision to utilize a power bank as the Raspberry Pi's power source in order to give clients mobility. The Raspberry Pi, a popular single-board computer [30], serves this purpose. OpenCV on a Raspberry Pi allows for the incorporation of all the major techniques and operations for image processing. We are using a 32 GB, class 10 SD card in our Raspberry Pi. Also, because the Raspberry Pi camera's cord is clumsy and unwieldy, we are utilising a USB camera instead.

3.3 YOLO:

The fundamental needs of our system are satisfied by YOLO, a real-time multi-object identification approach that is very quick

International Journal of Science and Engineering Applications
Volume 14-Issue 01, 34 – 42, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1401.1009

[31]. YOLO uses a single convolutional neural network to analyse an entire image, segmenting it into a S x S grid and then building boundary boxes around each segment to predict the chance of identifying, localising, and authenticating objects inside it is shown in Fig.2.

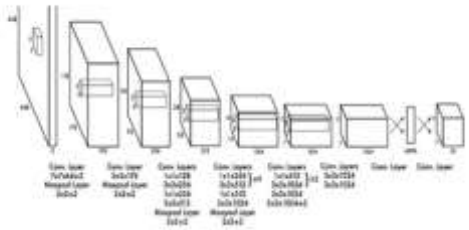


Fig.2. The YOLO Model's convolutional neural network.

YOLO expects several bounding boxes to exist in each grid cell. The most prominent union intersection (IOU) with the truth was selected for this research [32]. Specialised applications arise from the bounding box assumptions that underlie this method. The precision with which one can foretell certain dimensions and aspect ratios improves with each successive projection. In order to calculate the loss, YOLO adds up the discrepancy between the predicted and observed values. The network architecture of YOLO (You Only Look Once) comprises multiple components that facilitate efficient and precise object detection. The initial YOLO architecture, referred to as YOLOv1, pioneered the approach of considering object detection as a regression problem. In this approach, the neural network directly estimates the coordinates of the bounding boxes and the probabilities of different object classes. The neural network receives an input image that is partitioned into a grid of fixed dimensions at the input layer.

The YOLOv1 model commonly employs a sequence of convolutional layers to extract relevant features from the input image. These layers utilise filters of small dimensions, such as 3x3, and implement non-linear activation functions such as the Rectified Linear Unit (ReLU). Convolutional layers are typically succeeded by downsampling layers, such as max-pooling, that effectively decrease the spatial dimensions of the feature maps while simultaneously augmenting their depth. The process of downsampling aids in capturing features at varying scales and abstract levels. Following the downsampling process, it is common for the network to include fully connected layers in order to effectively handle the spatial information and extract features at a higher level.

The ultimate layer of the network is tasked with generating predictions. The system comprises a collection of neurons that generate the bounding box coordinates and class probabilities for every grid cell. The predictions consist of the coordinates of the bounding box (x, y, width, height) and the confidence score indicating the likelihood of an object being present. The class probabilities indicate the probability of various object categories within the given bounding box. During the training process, the YOLO algorithm optimises the network parameters by utilising a loss function that merges both the localization loss, which pertains to the precision of predicted bounding boxes, and the classification loss, which pertains to the accuracy of predicted class probabilities. Since the inception of the original YOLOv1, subsequent iterations including YOLOv2[25], YOLOv3[26], and YOLOv4 [27] have been introduced, incorporating various enhancements aimed at enhancing both accuracy and speed. The proposed enhancements encompass the incorporation of supplementary convolutional layers, the implementation of feature extraction at various scales, and the integration of skip connections to enhance feature fusion. The network architecture of YOLO is specifically designed to accomplish real-time object detection. This is achieved by efficiently processing the entire image in a single pass and directly predicting bounding boxes and class probabilities. OPENCV stands for "Open-source computer vision," and it is a collection of programmes optimised for use in real-time imaging [28]. It is possible that the toolkit has more than 2500 efficient techniques. Camera tracking, object recognition, character descriptions, and more are all possible applications of these techniques.

4. RESULT AND DISCUSSION

Fig.3. Custom Object Detection for Parking with YOLO

In Fig.3., the method is collecting a person, a bus, a handbag, and a backpack and identifying them based on how they seem. Fig.3 illustrates the proficient identification of targets within a pre-existing image through the utilisation of the YOLO algorithm. The algorithm has undergone training using the COCO dataset, a widely utilised dataset for object detection. It effectively and accurately identifies all objects depicted in the image. The algorithm showcases real-time activity recognition in addition to its object detection capabilities. This implies that the system has the capability to not only identify objects, but also discern various activities or actions occurring within the observed environment.

In addition, the algorithm offers confidential information in the process of classifying data. This suggests that it has the capability to provide more precise details regarding the identified objects, based on their specific class or category. In terms of performance, the system demonstrates a frame rate of 4 to 6 frames per second when executed on a CPU-based system. This implies that the system has the capability to efficiently process and analyse a range of 4 to 6 images per second in real-time. It is essential to acknowledge that the speed of object detection may vary based on the hardware employed and the optimisations applied. To enhance the speed of identification and recognition processes, it is advisable to utilise a system that is based on GPU technology. Graphics Processing Units (GPUs) are renowned for their exceptional parallel processing capabilities, enabling remarkable acceleration of computationally-demanding tasks such as object detection [30]. The utilisation of a GPU-based system

facilitates a more expedited identification and recognition process, thereby enhancing real-time performance.

Fig.4. Real-time room object detection

The algorithm has correctly recognised the monitor, book, mouse, cell phone, cup, and potted plant are in Fig.4. This was drawn from a real-world incident.

Fig.5. Detection of objects in a classroom in real time

The system had no trouble identifying the chair and the individual in the real-world situation, is shown in Fig.5. This sight at the bus stop is real as shown in Fig.6. This image can show a person, an automobile, or a handbag. This is a real-world office space arrangement. This image shows a person, a potted plant, a laptop, a dining table, a book, and a chair. The images are some real time scenarios that we take to test our system. In this system we used YOLO Version 3 algorithm [31]. Using this version, we get an average accuracy of about 98 percentage of each of the objects in the images. Fig.6. Bus halt object detection in real time

Fig.6. Bus halt object detection in real time

The main objects that are detected in those images are Person, Football, Book, Laptop, Potted plant, Cup, etc [32] – [35]

Fig.7. Real-time item detection in the room

This scene depicts a real room. This picture includes a dining table, chair, bed, cup, and bottles.

Fig.8. An actual football field's turf is detected in real time.

International Journal of Science and Engineering Applications
Volume 14-Issue 01, 34 – 42, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1401.1009

This is an actual football pitch scene. Football and a person may be seen in this image.

Fig.9. Office item identification in real time

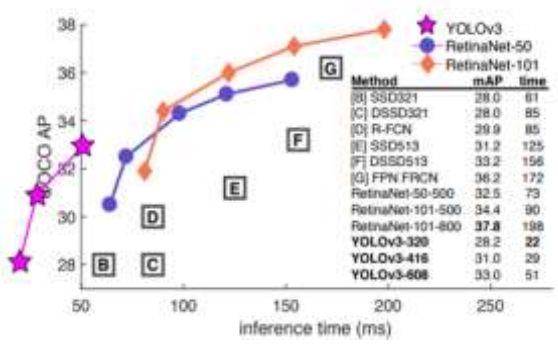
There are Further versions of YOLO and they give more accuracy and speed than this version. But even with this version we got a satisfying accuracy. Table 1 presents the outcomes of object detection utilising the YOLO algorithm on different figures. Each row within the table corresponds to a distinct figure, while the columns present details pertaining to the identified classes, accuracy percentages, and frame rates. Fig.3 illustrates the performance of the YOLO algorithm in object detection, specifically in identifying various objects such as individuals, buses, handbags, and backpacks. The algorithm demonstrates a remarkable level of accuracy, reaching 99%, in effectively detecting individuals and buses [36] – [39]. However, the accuracy decreases to 66% for handbags and 50% for backpacks. The frame rate for this figure is 445.20 milliseconds (ms). According to Fig.4, the algorithm successfully identifies various objects such as TV monitors, mice, books, cups, cellphones, potted plants, and vases. The system demonstrates a notable level of precision, achieving an accuracy rate of 96% for TV monitors and a perfect accuracy rate of 100% for mice. The detection of other objects exhibits varying levels of accuracy. The frame rate for this figure is 478.62 milliseconds. Fig.5 is dedicated to the detection of individuals and automobiles. The algorithm demonstrates a moderate level of accuracy, achieving a 57% success rate in detecting individuals and a high level of accuracy, reaching 97%, in detecting vehicles. The frame rate for this figure is 378.62 milliseconds. Fig.6 illustrates the algorithm's efficacy in detecting individuals, automobiles, and purses. The system demonstrates a remarkable level of precision, achieving a 100% accuracy in detecting individuals. However, it exhibits a comparatively lower accuracy rate of 37% for identifying cars and 44% for recognising handbags. The frame rate for this figure is 578.62 milliseconds. According to the findings presented in Fig.7, the algorithm successfully identifies various objects including beds, chairs, dining tables, cups, and bottles. The model demonstrates a commendable level of precision, accurately detecting beds with a success rate of 93%. It also exhibits varying degrees of accuracy when identifying other objects. The frame rate for this figure is 449.51 milliseconds. The primary objective of Fig.8 is to accurately identify and detect individuals as well as sports balls. The algorithm demonstrates a remarkable level of accuracy, achieving a perfect detection rate of 100% for identifying individuals and a notably high accuracy rate of 95% for detecting sports balls. The frame rate for this figure is 352.38 milliseconds. Finally, Fig.9 showcases the algorithm's efficacy in accurately detecting various objects such as laptops, books, individuals, potted plants, and chairs. The model demonstrates a notable level of accuracy when detecting laptops and individuals, while exhibiting comparatively lower accuracy when identifying other objects. The frame rate for this figure is 459.08 milliseconds. The table presents an analysis of the performance of the YOLO algorithm in detecting different objects across various images. The report showcases the level of accuracy attained for each object class and the frame rate at which the algorithm functions. These findings enhance the comprehension of the algorithm's efficacy in object detection tasks and can serve as a valuable point of reference for researchers and practitioners in the field of computer vision.

Fig.10. This graphic has been modified from the Focal Loss study [29].

Fig.10 represents a modified graphical representation that has been derived from a comprehensive study conducted on Focal Loss. The Focal Loss is a specialised loss function that is frequently employed in object detection tasks to effectively tackle the challenge of class imbalance. YOLOv3 operates far more quickly than competing detection techniques of equivalent capability**.**

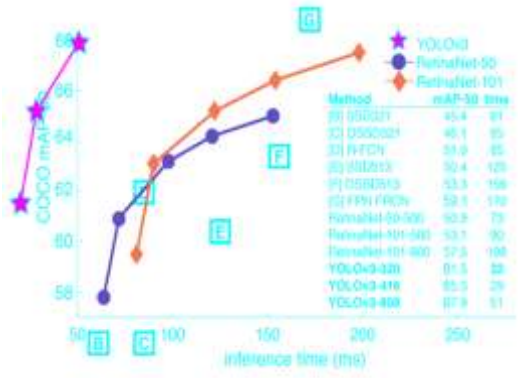
Figures	Class	Accuracy	Fps (ms)
Fig.3	person	99%	445.20
	Bus	99%	
	Handbag	66%	
	backpack	50%	
Fig.4	Tv monitor	96%	478.62
	Mouse	100%	
	Book	94%	
	Cup	98%	
	Cellphone	60%	
	pottedplant	91%	
	vase	51%	
Fig.5	Person	57%	378.62
	Car	97%	
Fig.6	Person	100%	578.62
	Car	37%	
	Handbag	44%	
Fig.7	Bed	93%	449.51
	Chair	85%	
	Dining Table	44%	
	Cup	98%	
	Bottle	60%	
Fig.8	person	100%	352.38
	sports ball	95%	
Fig.9	Laptop	99%	459.08
	Book	35%	
	person	100%	
	pottedplant	84%	
	chair	99%	

Table 1: Class Accuracy and Frames per Sec. of different Figures.



International Journal of Science and Engineering Applications
Volume 14-Issue 01, 34 – 42, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1401.1009

The modified graphic emphasises that YOLOv3, a particular iteration of the YOLO algorithm, demonstrates superior speed in comparison to other competing detection techniques that provide similar functionalities[41]-[43]. YOLOv3 is renowned for its capability to conduct real-time object detection, enabling rapid and efficient processing and identification of objects. The focal point lies in the fact that YOLOv3 not only attains exceptional accuracy in object detection, but also accomplishes this feat with remarkable speed. The speed advantage of YOLOv3 distinguishes it from other competing techniques, positioning it as the preferred choice in scenarios where real-time object detection is crucial



.

Fig.11. Garph of Modified YOLOv3 [40], here presenting the speed/accuracy trade-off on the mAP at .7 IOU metric.

The graph is positioned at a relatively high level and towards the left side. This indicates that the algorithm, most likely YOLOv3, demonstrates outstanding accuracy performance. The graph depicts a positive correlation between the position and the mAP score, suggesting that higher precision and recall in object detection are associated with higher values on the graph. Furthermore, the algorithm's placement on the far left suggests that it attains a notable level of accuracy while also operating at a commendable speed. According to the findings, YOLOv3 demonstrates a commendable equilibrium between accuracy and speed, surpassing alternative techniques in both these aspects. However, in the absence of specific details from Fig.11, such as axis labels or specific data points, it becomes difficult to offer a more accurate explanation. Based on the information provided, it can be inferred that the modified graph illustrates the notable accuracy and speed of YOLOv3, thereby highlighting its efficacy as an object detection algorithm.

5. CONCLUSION

Many methods have been developed in recent years to aid the visually impaired in recognising objects in their surroundings, but none of them are completely satisfactory. Our mission is to make it possible for the visually impaired to have access to a reliable and comfortable item detecting system. To save consumers needing to take pre-cut pictures, we utilise a USB camera to take pictures while they use our service. In this scenario, deep learning and the YOLO feature extraction technique are used. The YOLO fix helps with object recognition by picking the whole picture in one go, slicing it up into grids, then guesstimating the grid cells' bounding box coordinates and class labels. Sing YOLO's top quality is how quickly it can be sung. It is lightning fast and can handle a wide variety of objects with ease. This gadget uses text-to-speech technology to provide audible explanations of the user's surroundings, thereby rendering the visually handicapped invisible and facilitating their freedom of movement. The suggested approach is mobile, dependable, and cost-effective. In addition to providing a safe and immersive virtual environment, this technology also offers peace of mind by revealing the name of the recognised item.This project's long-term goals include enhancing computer vision rates, which can be handled by employing the Python library, and supplying an accurate estimate of the range between individuals and objects [44]. You should, however, take speedier hardware into perspective if you are creating an application that uses a lot of quickly changing objects. Furthermore, the same method may be used to do both text and facial recognition. The system will however be mostly comparable.

Conflict of Interest: Authors declare that there is no conflict of interest.

5. 6. REFERENCES

[1] Tielsch, J. M., Sommer, A., Witt, K., Katz, J., & Royall, R. M. (1990). Blindness and visual impairment in an American urban population: the Baltimore Eye Survey. Archives of ophthalmology, 108(2), 286-290.

[2] Ntakolia, C., Dimas, G., & Iakovidis, D. K. (2022). User-centered system design for assisted navigation of visually impaired individuals in outdoor cultural environments. Universal Access in the Information Society, 1-26.

[3] Lilley, J. M., Arie, T., & Chilvers, C. E. D. (1995). Accidents involving older people: a review of the literature. Age and Ageing, 24(4), 347-367.

[4] Parkin, J., & Smithies, N. (2012). Accounting for the needs of blind and visually impaired people in public realm design. Journal of urban design, 17(1), 135-149.

[5] Bahadir, S. K., Koncar, V., & Kalaoglu, F. (2012). Wearable obstacle detection system fully integrated to textile structures for visually impaired people. Sensors and Actuators A: Physical, 179, 297-311.

[6] babu Nuthalapati, S. Advancements in Generative AI: Applications and Challenges in the Modern Era. (2024). International Journal of Science and Engineering Applications

Volume 13-Issue 08, 106 – 111, DOI: 10.7753/IJSEA1308.1023

[7] Chan, M., Estève, D., Escriba, C., & Campo, E. (2008). A review of smart homes—Present state and future challenges. Computer methods and programs in biomedicine, 91(1), 55-81.

[8] Md. Atikur Rahman, Muhammad Sheikh Sadi (2021) IoT Enabled Automated Object Recognition for the Visually Impaired, Computer Methods and Programs in Biomedicine Update, Volume 1, 100015, ISSN 2666-9900.

[9] Soman, S. P., Kumar, G. S., Nuthalapati, S. B., Zafar, S., & Abubeker, K. M. (2024). Internet of things assisted deep learning enabled driver drowsiness monitoring and alert system using CNN-LSTM framework. Engineering Research Express, 6(4), 045239.

[10] Mahendru, Mansi, and Sanjay Kumar Dubey (2021) Real time object detection with audio feedback using Yolo vs. In 2021 11th International Conference on Cloud Computing, pp. 734-740. IEEE, 2021.

[11] Ferdousi Rahman, I. J. R., Farhin, N., Uddin, J.(2019). An assistive model for visually impaired people using yolo and mtcnn.

[12] babu Nuthalapati, S. (2023). AI-enhanced detection and mitigation of cybersecurity threats in digital banking. Educ. Adm. Theory Pract., 29(1), 357-368.

[13] Jiang, R., Lin, Q., Qu, S. (2016). Let blind people see: real-time visual recognition with results converted to 3D audio.Report No. 218, Standord University, Stanford, USA.

[14] Dongre, A. (2020). Visual Recognition Based System To Assist Blind Persons (Doctoral dissertation, Dublin, National College of Ireland).

[15] Nuthalapati, S. B., & Nuthalapati, A. (2024). Transforming Healthcare Delivery via IoT-Driven Big Data Analytics in A Cloud-Based Platform. Journal of Population Therapeutics and Clinical Pharmacology, 31(6), 2559-2569.

[16] 16. H. Jabnoun, F. Benzarti and H. Amiri, (2015) Object detection and identification for blind people in video scene, 15th International Conference on Intelligent Systems Design and Applications (ISDA), pp. 363-367, doi:10.1109/ISDA.2015.7489256.

[17] Yingli Tian (2020). Visual Recognition Based System To Assist Blind Persons (Doctoral dissertation, Dublin, National College of Ireland).

[18] Ashiq, F., Asif, M., Ahmad, M. B., Zafar, S., Masood, K., Mahmood, T., ... & Lee, I. H. (2022). CNN-based object recognition and tracking system to assist visually impaired people. IEEE Access, 10, 14819-14834.

[19] Babu Nuthalapati, S., & Nuthalapati, A. (2024). Accurate weather forecasting with dominant gradient boosting using machine learning. Int. J. Sci. Res. Arch, 12(2), 408-422.

[20] Yohannes, E., Lin, P., Lin, C. Y., & Shih, T. K. (2020, December). Robot eye: automatic object detection and recognition using deep attention network to assist blind people. In 2020 International Conference on Pervasive Artificial Intelligence (ICPAI) (pp. 152-157). IEEE.

[21] Joshi, R. C., Yadav, S., Dutta, M. K., & Travieso-Gonzalez, C. M. (2020). Efficient multi-object detection and smart navigation using artificial intelligence for visually impaired people. Entropy, 22(9), 941.

[22] Nuthalapati, A., Abubeker, K. M., & Bushara, A. R. (2024, September). Internet of Things and Cloud Assisted LoRaWAN Enabled Real-Time Water Quality Monitoring Framework for Urban and Metropolitan Cities. In 2024 IEEE North Karnataka Subsection Flagship International Conference (NKCon) (pp. 1-6). IEEE.

[23] Kadhim, M. R., & Oleiwi, B. K. (2022). Blind assistive system based on real time object recognition using machine learning. Engineering and Technology Journal, 40(1), 159-165.

[24] Bushara, A. R., RS, V. K., & Kumar, S. S. (2023). Classification of Benign and Malignancy in Lung Cancer Using Capsule Networks with Dynamic Routing Algorithm on Computed Tomography Images. Journal of Artificial Intelligence and Technology.

[25] Devi, S. K., & Subalalitha, C. N. (2022). Deep learning-based audio assistive system for visually impaired people. CMC-COMPUTERS MATERIALS & CONTINUA, 71(1), 1205-1219.

[26] Nuthalapati, A. (2024). Cloud data center performance optimization through machine learning-based workload forecasting and energy efficiency.

[27] 27. Hersh, Marion. "Route learning by blind and partially sighted people." Journal of Blindness Innovation and Research 10, no. 2 (2020).

[28] Chen, R. C., Saravanarajan, V. S., & Hung, H. T. (2021). Monitoring the behaviours of pet cat based on YOLO model and raspberry Pi. International Journal of Applied Science and Engineering, 18(5), 1-12.

[29] Islam, R. B., Akhter, S., Iqbal, F., Rahman, M. S. U., & Khan, R. (2023). Deep learning based object detection and surrounding environment description for visually impaired people. Heliyon, 9(6).

[30] Nuthalapati, A. (2024). Architecting data lake-houses in the cloud: Best practices and future directions. Int. J. Sci. Res. Arch, 12(2), 1902-1909.

[31] Pujara, A., & Bhamare, M. (2022, November). DeepSORT: Real Time & Multi-Object Detection and Tracking with YOLO and TensorFlow. In 2022 International Conference on Augmented Intelligence and Sustainable Systems (ICAISS) (pp. 456-460). IEEE.

[32] Alsanad, H. R., Sadik, A. Z., Ucan, O. N., Ilyas, M., & Bayat, O. (2022). YOLO-V3 based real-time drone detection algorithm. Multimedia tools and applications, 81(18), 26185-26198.

[33] AR, B., RS, V. K., & SS, K. (2023). LCD-capsule network for the detection and classification of lung cancer on computed tomography images. Multimedia Tools and Applications, 82(24), 37573-37592.

[34] Han, X., Chang, J., & Wang, K. (2021). Real-time object detection based on YOLO-v2 for tiny vehicle object. Procedia Computer Science, 183, 61-72.

[35] Wang, K., & Liu, M. (2022). YOLOv3-MT: A YOLOv3 using multi-target tracking for vehicle visual detection. Applied Intelligence, 52(2), 2070-2091.

[36] Bushara, A. R., RS, V. K., & Kumar, S. S. (2024). The Implications of Varying Batch-Size in the Classification of Patch-Based Lung Nodules Using Convolutional Neural Network Architecture on Computed Tomography Images. Journal of Biomedical Photonics & Engineering, 10(1), 39-47,.

[37] Zhang, X., & Wang, G. (2022). Stud pose detection based on photometric stereo and lightweight YOLOv4. Journal of Artificial Intelligence and Technology, 2(1), 32-37.

[38] Kohut, P. (2013). Mechatronics systems supported by vision techniques. Solid State Phenomena, 196, 62-73.

[39] Subeh, P., & Bushara, A. R. (2024). Cloud data centers and networks: Applications and optimization techniques. International Journal of Science and Research Archive, 2024, 13 (02), 218-226.

[40] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar. Focal loss for dense object detection. arXiv preprint arXiv:1708.02002, 2017.

[41] Akenine-Moller, T., & Strom, J. (2008). Graphics processing units for handhelds. Proceedings of the IEEE, 96(5), 779-789.

[42] Kashika, P. H., & Venkatapur, R. B. (2022). Automatic tracking of objects using improvised Yolov3 algorithm and alarm human activities in case of anomalies. International Journal of Information Technology, 14(6), 2885-2891.

[43] Kheder, M. Q., & Ali, M. A. (2022). IoT-Based Vision Techniques in Autonomous Driving: A Review. ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal, 11(3), 367-394.

[44] Muhammed Kunju, A. K., Baskar, S., Zafar, S., & AR, B. (2024). A transformer based real-time photo captioning framework for visually impaired people with visual attention. Multimedia Tools and Applications, 1-20.

International Journal of Science and Engineering Applications
Volume 14-Issue 01, 43 – 46, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1401.1010

Multi-Point Temperature Detection System Design

Kangning Song
School of Electronic Information and Electrical Engineering
Yangtze University
Jingzhou, China

Abstract: In the context of the rapid development of the beekeeping industry, accurately measuring the temperature inside the hive is crucial for beekeeping. Traditional methods of measuring the temperature inside hives consume a significant amount of manpower and resources. This paper designs a wireless temperature measurement system for remote temperature measurement inside hives based on the NRF905 wireless RF chip. The system designed in this paper includes both a host computer system and a slave computer system, which can be divided into six modules: power supply module, temperature measurement module, minimum system, NRF905 wireless RF module, display module, and over-temperature alarm module. This design allows for the remote display of temperature data inside the hive, enabling real-time monitoring of the bee growth environment. By combining wireless communication technology with agricultural production, this design provides a relatively good solution for hive temperature detection, which is of great significance for promoting the modernization of beekeeping.
Keywords: Wireless communication; DS18B20; NRF905;

1. INTRODUCTION

Temperature, as one of the most common physical quantities in life, not only affects people's daily lives but is also closely related to industrial and agricultural production. In agricultural production, temperature influences the yield of crops and the growth conditions of livestock. Beekeeping, as one of the components of agriculture, also has high requirements for temperature, thus necessitating the design of a temperature detection system to monitor the temperature inside hives in a timely manner. With the rapid development of temperature measurement technology, temperature sensors have been widely used in various temperature measurement and detection systems. This has led to a significant increase in signal cables, which not only complicates wiring and increases the workload of installation but also poses the risk of electromagnetic interference, affecting the accuracy of the measured temperature data, and making maintenance difficult. To overcome these drawbacks, there is an urgent need to develop wireless sensing technology that meets data communication requirements, in order to address the issues caused by connecting cables and reduce production costs.

Using wireless technology for temperature data collection and measurement offers several advantages, such as flexibility, ease of maintenance, and low cost. This is particularly beneficial in harsh environmental conditions where it is difficult to lay cables. By utilizing wireless technology to transmit sensor data, the costs of equipment and labor are reduced, transmission reliability is improved, and the system can meet most usage requirements. It can be widely promoted in the context of beekeeping applications.

This paper focuses on the measurement of temperature inside bee hives, designing a temperature monitoring system capable of remote real-time monitoring. The design integrates wireless communication technology with agricultural production. Specifically, for the task of measuring temperature inside bee hives, a small-volume, high-stability, easy-to-maintain, and reliable data transmission wireless temperature data transmission system is designed, consisting of temperature sensors, microcontrollers, RF chips, and other components. This system enables the measurement, wireless transmission, and display of temperature at multiple points inside the hive, allowing for real-time remote monitoring of the hive's temperature. It provides technical support for the production and life of bees and for increasing honey yield[1].

2. OVERALL SYSTEM DESIGN

The design, differentiated by functionality, is divided into six modules: power supply module, temperature measurement module, minimum system, NRF905 RF module, display module, and over-temperature alarm module.

The overall layout of the design is split into two parts: the upper computer (host) and the lower computer (slave).The lower computer mainly consists of four parts: the minimum system, temperature measurement module, power supply module, and NRF905 RF module. It is responsible for accurately measuring temperature data from multiple points and sending this temperature data to the upper computer via the NRF905 RF module.The upper computer receives the temperature data sent by the lower computer through the NRF905 RF module and displays it using the display module. If the temperature exceeds a threshold, it triggers an over-temperature alarm. Communication between the upper and lower computers is facilitated by wireless transceiver chips, together forming a wireless temperature monitoring system[2].

This design utilizes the digital temperature sensor DS18B20, which can directly output a 9 to 12-bit digital temperature value. Eight DS18B20 sensors are used to measure temperature at eight different points inside the beehive. The system design employs the NRF905 as the wireless communication chip, based on two STC51 microcontroller units. The temperature data collected by the temperature sensors is sent through the SPI interface to the NRF905 transmitter and then to the NRF905 receiver. After the receiver gets the temperature data, it is displayed on the display module, enabling remote monitoring of multiple points within the beehive.

3. SYSTEM HARDWARE CIRCUIT DESIG

3.1 Minimum system

The AT89C51 is a microcontroller produced by ATMEL Corporation. It includes an 8-bit CPU and is an 8-bit CMOS microcontroller. The chip contains 8KB of programmable FLASH memory, which offers strong compatibility and makes

International Journal of Science and Engineering Applications
Volume 14-Issue 01, 43 – 46, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1401.1010

debugging and development very convenient. Additionally, it has excellent expandability, allowing it to work with other external chips to achieve functions such as button control, analog-to-digital conversion, and display control. The minimum system of a microcontroller refers to the simplest working environment in which the microcontroller can execute programs normally. An AT89C51 connected externally to a clock circuit, a reset circuit, and basic input modules constitutes a minimum system.

3.2 Display module

The LCD1602 display is a character-type display device, consisting of 32 blocks of 5×8 dot matrices. Each character requires only one complete dot matrix block, allowing the device to display up to 32 characters simultaneously. The display operates on a supply voltage of +5V. The LCD1602 can interface with a microcontroller using either an 8-bit data bus or a 4-bit data bus.

3.3 Temperature measurement module

The DS18B20 is a common digital temperature sensor, known for its small size, strong interference resistance, and high precision. It uses a one-wire interface, requiring only a single data line for communication with the controller. Each DS18B20 has a unique serial number, which allows for multi-point networking and effectively reduces the occupation of microcontroller input/output interfaces[3]. Additionally, the device can be powered using a parasitic power supply, drawing energy from the communication line. This offers advantages such as good economy, strong interference resistance, flexible power supply, and high measurement accuracy.

3.4 NRF905 wireless RF module

This design selects the NRF905 wireless transceiver chip as the core of wireless communication. This device is an independent transceiver introduced by the Norwegian company NORDIC, operating normally within a voltage range of 1.9~3.6V. It can perform wireless communication on three ISM bands: 433MHz, 868MHz, and 915MHz[4]. When using the 433MHz communication frequency, which offers the longest transmission distance and the least signal attenuation, the communication range can reach up to 500 meters, with a packet loss rate of less than 1%. Additionally, it is not limited by communication networks, overcoming the risks of GPRS disconnection and the short communication range and high packet loss rate associated with Wi-Fi and Bluetooth technologies [5]. The channel switching speed is fast, capable of making a switch operation within 650μs. It supports the low-power ShockBurst™ transmission mode, which automatically handles the preamble and CRC checksum, thereby simplifying the tasks of the microcontroller and improving the efficiency of data transmission.

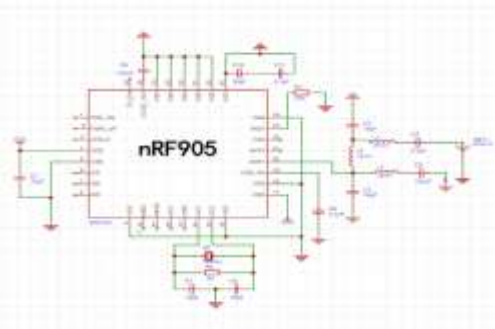


Figure. 1 NRF905 Pinout Diagram

4. SOFTWARE PROGRAM DESIGN

The software part mainly involves the writing of the host and slave programs, requiring modular programming. The slave system collects temperature values from each temperature sensor at various measurement points and sends them to the NRF905 wireless RF module, which then wirelessly transmits the data to the host's NRF905 wireless RF module, completing the transmission process. The slave program includes the main program, temperature collection program, and wireless transmission program.

The host system receives data transmitted from the nodes through the NRF905 wireless RF module, sends it to the microcontroller, and displays it on the LCD1602 screen, while also using an LED for over-temperature alarm. The host program includes the main program, wireless reception program, data display program, and over-temperature alarm program.

4.1 Temperature acquisition program

In the temperature acquisition module of this design, the host is an AT89C51 microcontroller, and the slaves are eight DS18B20 sensors connected in parallel on the bus. When the AT89C51 needs to receive temperature data measured by the DS18B20, it first actively initiates an initialization sequence to initialize the device, then enters the write sequence to send temperature conversion commands, ROM match commands, and read memory commands. Finally, the AT89C51 initiates a read sequence to retrieve the temperature data.

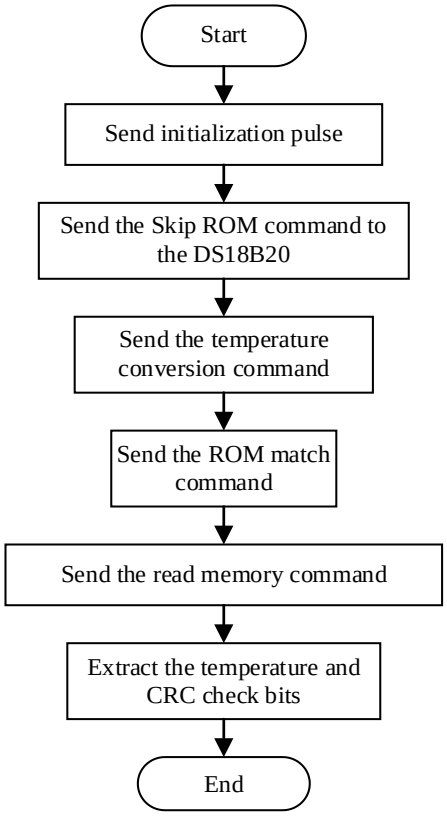


Figure. 2 Temperature Measurement Program Flowchart

International Journal of Science and Engineering Applications
Volume 14-Issue 01, 43 – 46, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1401.1010

4.2 Wireless transmission program

The NRF905 temperature data transmission process is divided into the following steps:

A. When the AT89C51 in the temperature measurement and transmission module receives temperature data, the AT89C51 first sets the SPI interface rate and simultaneously transmits the address of the NRF905 in the temperature reception and display module and the temperature data to the NRF905 through the SPI interface.

B. The AT89C51 controls TRX_CE and TX_EN to go high, setting the NRF905 to ShockBurst™ transmission mode.

C. The NRF905 enters the ShockBurst™ transmission state:

1. The RF configuration register is automatically enabled.
2. The NRF905 automatically combines the preamble and CRC checksum with the temperature data and packs them together.
3. The data packet is transmitted.
4. After the data packet is sent to the NRF905 receiver in the temperature reception and display module, the data ready pin DR automatically goes high.

D. If the automatic retransmission parameter AUTO_RETRAN is set high during device initialization, the NRF905 continuously repeats the data packet transmission.

E. When TRX_CE is set low, the transmission process is completed, the ShockBurst™ transmission mode ends, and it automatically enters the standby power-saving mode.

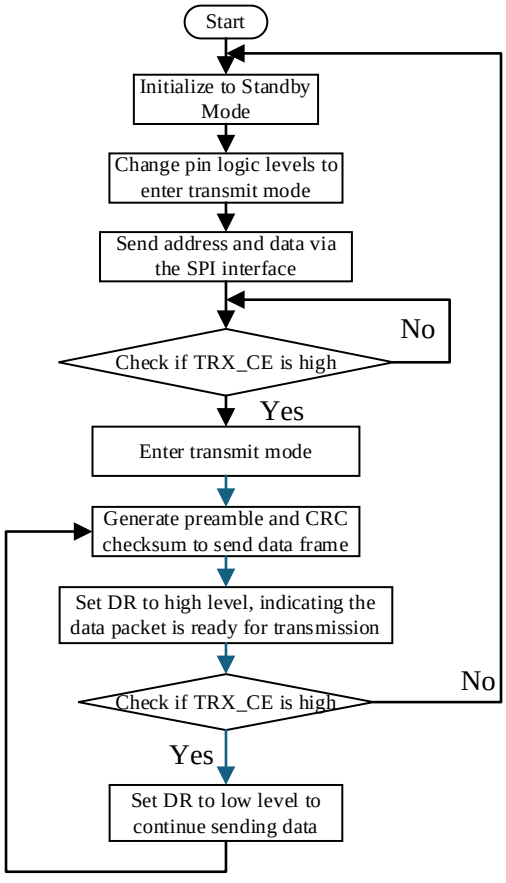


Figure.3 Wireless Transmission Program Flowchart

4.3 Wireless reception program

The NRF905 temperature data reception process is divided into the following steps:

A. The AT89C51 sets TRX_CE to a high level and TX_EN to a low level, configuring the NRF905 to ShockBurst™ reception mode.

B. After a brief 650μs preparation time, the NRF905 continuously monitors the external carrier signal.

C. When the NRF905 detects a carrier signal at 433MHz, the carrier detect pin CD is set to a high level.

D. The address in the data packet is checked to see if it matches the address set in the internal register of the NRF905 in the temperature reception and display module during initialization. If they match, the address match pin AM is set to a high level.

E. The preamble synchronization sequence and CRC check are used to verify the accuracy of the temperature data. Once the data is confirmed to be correct, the NRF905 automatically removes the preamble, address, and CRC check bits from the data packet and sets the data ready pin DR to a high level.

F. The AT89C51 sets TRX_CE to a low level, completing the reception process. The ShockBurst™ reception mode ends, and it automatically enters the standby power-saving mode.

G. The AT89C51 in the temperature reception and display module transfers the data to the AT89C51 via the SPI interface at the user-defined SPI interface rate.

H. After the complete transfer of data to the AT89C51, the NRF905 in the temperature reception and display module sets the data ready pin DR and the address match pin AM to low.

I. At this point, the NRF905 can enter any operating mode or power-down power-saving mode.

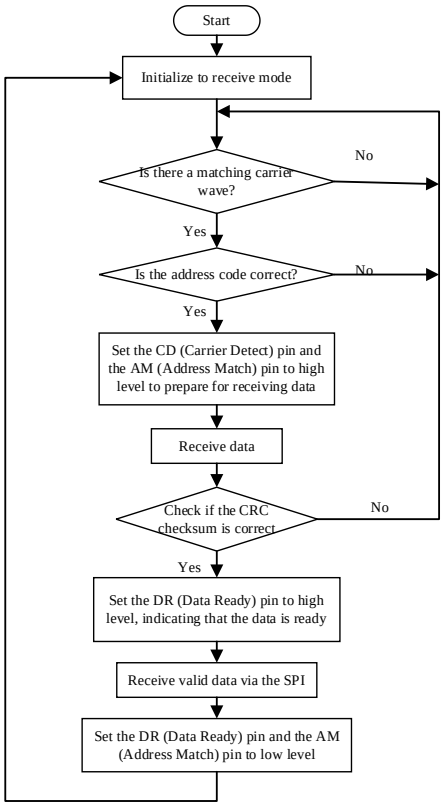


Figure.4 Wireless Reception Program Flowchart

5. CIRCUIT HARDW ARE ASSEMBLY AND SYSTEM DEBUGGING

This design uses Keil uVision5 for programming and Proteus8.6 for circuit diagram simulation. Keil is one of the most commonly used software development platforms for microcontroller learners. It integrates a variety of functions such as editing, compiling, and simulation. In terms of programming languages, it not only supports assembly language, which is often used by beginners, but also supports the more concise and clear C language, catering well to all kinds of users. This has made it one of the most popular software for the development of programs for 8051 architecture microcontrollers.Proteus is a widely used circuit simulation software that not only covers circuit simulation and PCB design capabilities but also allows for the simulation of microcontroller circuits. It assists users in learning about and designing a variety of circuits. After adding the target file with the .hex extension, which has been successfully compiled by Keil, to the microcontroller, the two software programs operate in conjunction to simulate the data acquisition and display module. The simulation results show that the eight DS18B20 temperature sensors measure temperature and the LCD1602 displays the temperature normally, and it is possible to switch and display temperature data from different sensors using buttons. When the temperature exceeds the user-defined threshold set in the program, the temperature alarm light turns on, and the over-temperature alarm module functions properly.

6. REFERENCES

Based on the study of various short-range wireless communication technologies, this paper designs a short-range wireless temperature acquisition, transmission, reception, and display system that is based on the nRF905 RF transceiver module, DS18B20 digital temperature sensor, and the AT89C51 microcontroller.This paper addresses the need for multi-point temperature detection inside beehives by carrying out both hardware circuit design and relevant software programming. The hardware circuit design includes the selection of a microcontroller, temperature sensors, wireless transceiver chips, and display modules, as well as the circuit design for the temperature measurement module, wireless transmission module, and LCD display module. This paper employs partial simulation to simulate two important modules identified in the requirements. In terms of software program design, programs for temperature measurement and wireless transmission of temperature data have been developed.The design takes full advantage of the low power consumption characteristics of the selected components, significantly reducing the system's energy consumption. This allows the system to operate normally for extended periods in the field where power supply conditions are lacking, making it more suitable for beekeeping environments. The temperature measurement system is cost-effective, reliable in temperature measurement, has a high rate of accurate data transmission, is convenient for maintenance, and offers considerable economic benefits.

7. REFERENCES

[1] Hugo Hadjur, Doreid Ammar, Laurent Lefèvre, Toward a n intelligent and efficient beehive: A survey of precision beekeeping systems and services. Computers and Electro nics in Agriculture, 2022, Volume 192: 0168-1699.

[2] Wang, Xue Mei, and Guo Ping Li. “Granary Wireless Te mperature Monitoring and Alarm System Based on DS18 B20.” Applied Mechanics and Materials 220–223 (Nove mber 2012): 1625–27.

[3] Wu, Yan Xiang, Dan Liu, and Xing Hong Kuang. “A Te mperature Detecting System Based on DS18B20.” Advan ced Materials Research 328–330 (September 2011): 1806 –9.

[4] Huang, J., Liu, D., & Yuan, Q. (2019). An Anthurium Gr owth Environment Monitoring System Based on Wireless Sensor Network. International Journal of Online and Bio medical Engineering (iJOE), 15(05), pp. 69–85.

[5] Huang Tao, Fang Qing, Sun Qingye. Design of Unmanne d Boat Wetland Environmental Monitoring System Based on Microcontroller Technology [J]. Science, Technology and Innovation Application, 2024, 14(9): 50-55, 60.

International Journal of Science and Engineering Applications
Volume 14-Issue 01, 47 – 51, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1401.1011

Design of Temperature Smoke Alarm Based on Single Chip Computer

HaiJun Wang
School of Electronic Information and Electrical Engineering
Yangtze University
Jingzhou, China

Abstract: This paper presents an intelligent temperature and smoke monitoring and alarm system based on STC89C52 single chip computer. The system uses STC89C52 single chip microcomputer as the core controller, integrates DS18B20 temperature sensor and MQ-2 smoke sensor, and can continuously monitor the ambient temperature and smoke concentration in real time. In addition, the system is equipped with a Bluetooth communication module, which enables it to transmit the monitored temperature and smoke data to the mobile device in real time, thus improving the dynamic and interactive monitoring. When the monitoring data exceeds the preset safety threshold, the system will not only immediately start the audible and visual alarm to warn the field personnel to take prompt action, but also send the alarm information to the preset mobile device through the Bluetooth communication module to ensure that the relevant personnel can timely understand the situation and take the necessary preventive measures, enhancing the timeliness and effectiveness of the system. The simulation and real debugging results show that the system can run normally, which has a certain significance for environmental safety monitoring.

Keywords: Alarm system; STC89C52 MCU; Temperature sensor; Smoke sensor;

1. INTRODUCTION

Modern society attaches great importance to the safety of people's lives and property. With the acceleration of industrialization and urbanization, the residential density continues to increase, the complexity of office environment and living place increases, and the fire risk increases accordingly. It is very important to develop a temperature smoke alarm system with high precision, all-round monitoring ability and modern remote communication technology to detect fire hazards in advance and reduce fire accident losses.

In addition, with the continuous development of microelectronics technology, sensor technology and network communication technology, the performance of intelligent fire detection and early warning system has been greatly improved, which can more effectively strengthen fire safety management and ensure the safety of public life and property. In particular, the extensive application of single-chip technology makes electronic equipment more intelligent and miniaturized, and promotes the upgrading of fire detection system. Because of its advantages of low price, small size, low power consumption and high integration, SCM has become the preferred core control unit for designing modern temperature smoke alarms.[1]

This paper aims to design and develop a temperature smoke alarm system based on single chip microcomputer, which integrates high precision temperature and smoke monitoring technology and Bluetooth communication module. This design not only pursues high integration, strong dynamic monitoring ability and friendly user interaction, but also pays attention to cost efficiency, so as to meet the growing demand of intelligent alarm equipment in the market. By combining modern electronics and information technology, this project greatly improves the efficiency and intelligence level of fire detection, and broadens the application range of fire warning system. The innovation and promotion of this technology not only effectively improves the social security, reduces the casualties and property losses caused by fire, but also produces significant economic benefits, reduces the economic losses caused by fire, and contributes to social stability and the effective use of resources.

2. OVERALL SCHEME DESIGN

The overall scheme design is a key step in the development of temperature smoke alarms, which determines the core structure and function of the system. The scheme is designed around six main links of smoke monitoring, alarm, temperature monitoring, data display, Bluetooth communication and control keys to ensure the reliability and stability of the system. The design scheme integrates functional modules such as smoke monitoring, temperature monitoring, MCU control, character display and sound and light alarm into a complete system.

Temperature smoke alarm can be divided into single-chip microcomputer minimum system, temperature monitoring module, smoke concentration monitoring module, keying module, Bluetooth module, audible and visual alarm module and display module. Among them, the core control module is the smallest single-chip microcomputer system, which collects real-time information of external temperature and smoke concentration together with the smoke concentration monitoring module and temperature monitoring module, and then collaborates with other modules to complete threshold setting, sound and light alarm function, remote threshold setting, and real-time monitoring of external environment temperature and smoke concentration.

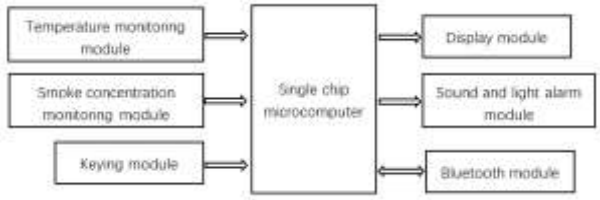


Figure. 1 System frame diagram

International Journal of Science and Engineering Applications
Volume 14-Issue 01, 47 – 51, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1401.1011

3. SYSTEM HARDWARE CIRCUIT

3.1 Minimal System of Microcontroller

The minimal system of a microcontroller refers to a complete and fully functional working environment for the microcontroller, including the core microcontroller and the necessary peripheral circuits, enabling it to operate normally and perform the designated tasks. In this system, the STC89C52 serves as the core component, providing control and computing capabilities for the entire system. The STC89C52 microcontroller, with its high cost-effectiveness, good compatibility, ease of programming and development, extensive application support, stable and reliable performance, as well as a rich set of peripheral interfaces, makes it an ideal choice for the control unit of temperature and smoke alarm systems. As an 8-bit microcontroller based on the 8051 core, the STC89C52 not only inherits the core features of the 8051, but also enhances processing power and memory capacity by increasing the operating frequency and expanding functionality.[2] Additionally, it supports online programming and serial downloading, features that make it widely used in numerous embedded system projects.

3.2 Smoke Concentration Monitoring Module

Smoke sensors are widely used in firefighting, security, and monitoring fields to detect the concentration of smoke in the environment, providing timely alerts to protect people and property. The MQ-2 sensor, due to its high sensitivity, broad detection range, and low cost, has become an ideal choice for smoke detection in fire alarm systems. It is capable of detecting a variety of combustible gases and smoke, making it particularly suitable for early fire detection. The working principle of the MQ-2 sensor is based on a tin dioxide (SnO2) semiconductor element, which changes its resistance when exposed to smoke or combustible gases. This change in resistance is converted into an electrical signal, which can be interfaced with a microcontroller, making it suitable for large-scale deployment and cost-effective applications.[3]

The core of the smoke monitoring module is the MQ-2 sensor, which is capable of real-time collection of smoke concentration data from the environment. The collected data is converted into voltage values at different concentrations using an ADC0832 analog-to-digital converter. Based on these voltage values, an ideal smoke concentration alarm threshold can be set, enabling real-time monitoring and alarm of smoke concentration..

3.3 Display Module Circuit Design

The core component of the display module is the 1602 LCD display, which is used to show the set threshold values as well as the monitored temperature and smoke concentration values from the external environment. The display circuit is shown in Figure 2.

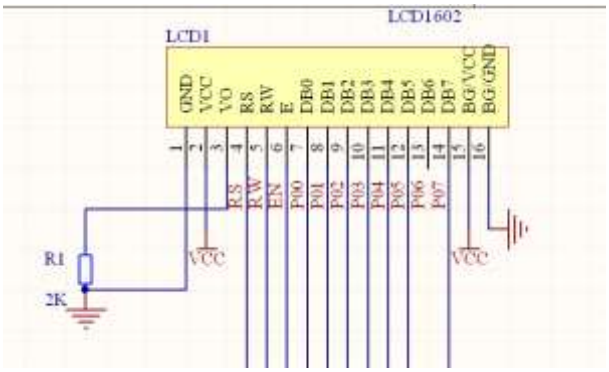


Figure .2 LCD1602 LCD

3.4 Sound and Light Alarm Module Circuit Design

In the circuit design, a resistor is placed in series between the base of the transistor and the microcontroller port, enabling control of the buzzer alarm and the on/off state of the LED. This configuration allows the microcontroller to directly control the transistor's conduction and cutoff by outputting signals, thereby achieving the switching function of the alarm or indicator light. The circuit diagram is shown in Figure 3.

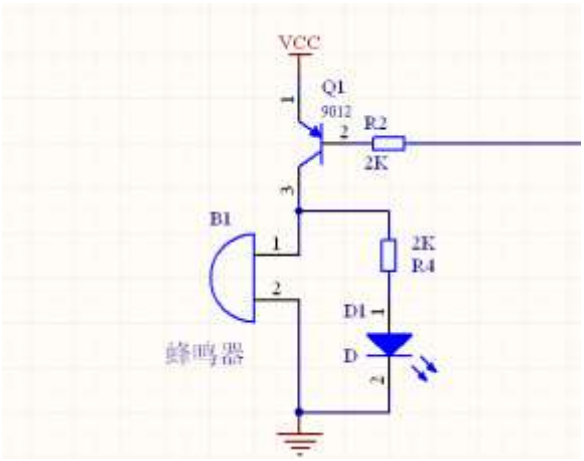


Figure.3 Audio alarm circuit diagram

3.5 Key Control Module Circuit Design

This module includes a set button, an alarm button, an increase button, and a decrease button. The set button allows the user to select whether to adjust the temperature threshold or the smoke concentration threshold. When the alarm button is pressed, the alarm will immediately trigger both sound and light warnings, which is crucial in emergency situations. The increase and decrease buttons are used to adjust the value of the threshold. The circuit diagram is shown in Figure 4.

International Journal of Science and Engineering Applications
Volume 14-Issue 01, 47 – 51, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1401.1011

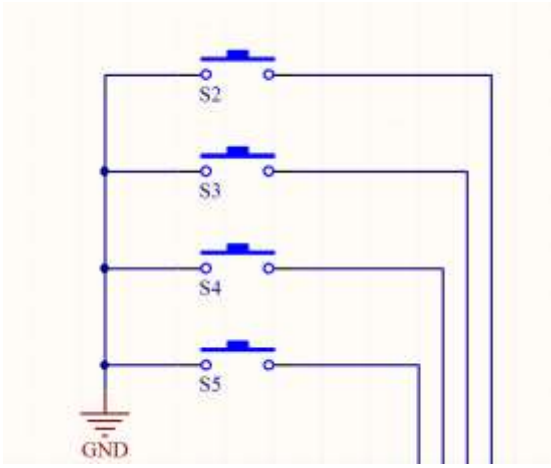


Figure.4 Key connection circuit diagram

3.6 Bluetooth module

Integrating a Bluetooth module (such as the HC-05 Bluetooth chip) into the system can significantly enhance the intelligence and convenience of the temperature and smoke alarm. It allows users to monitor the device status via a smartphone, perform remote operations (such as turning off the alarm or adjusting thresholds), and transmit and record data (such as temperature and smoke concentration). The HC-05 module offers excellent compatibility, a serial communication interface, and configurability, enabling wireless communication with microcontrollers and providing a better user experience and system scalability.[4]

The Vcc pin of the HC-05 is connected to the input power supply, the GND pin to ground, and the TX and RX pins of the HC-05 are connected to the P3.0 and P3.1 pins of the STC89C52, respectively, completing the serial communication between the Bluetooth module and the microprocessor.

3.7 Temperature monitoring module

The DS18B20 is chosen as the temperature sensor for the temperature monitoring module primarily because of its wide measurement range, high accuracy, and ability to directly interface with microcontrollers via a single-wire protocol. This simplifies system design while enhancing scalability. Its fast digital temperature conversion within 1 second and user-configurable temperature alarm settings make it particularly valuable in applications such as temperature control, industrial systems, and thermosensitive devices. The durability and stability of the DS18B20 are ensured even in complex environments, and its high cost-effectiveness makes it an ideal choice for cost-sensitive applications.

The interface circuit of the DS18B20 is very simple in design, mainly due to its unique single-wire communication method. In a typical DS18B20 interface circuit, the main components include the sensor itself, a pull-up resistor, the data line, and the microcontroller interface.[5] The data line is used both for data transmission and as a power supply line. The pull-up resistor plays a critical role in the DS18B20 interface circuit. The data line needs to be connected to a high voltage level through the pull-up resistor, which helps to ensure that the data line remains at a stable high level, allowing the DS18B20 to correctly receive and transmit signals. The circuit diagram is shown in Figure 5.

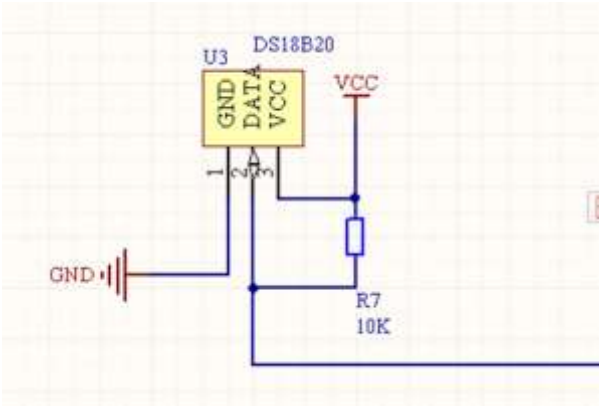


Figure.5 Circuit diagram of the temperature sensor port

4. SYSTEM SOFTWARE DESIGN

4.1 System Main Program Design and Flowchart

When the power is turned on and the switch is pressed to power the system, the system will initialize. During initialization, the temperature sensor and smoke sensor subroutines will be started to begin collecting environmental temperature and smoke concentration data. The system will transmit the collected data via the Bluetooth module to the connected device. Meanwhile, the collected temperature and smoke concentration data will be evaluated to check whether they exceed the preset thresholds. If the values exceed the thresholds, the system will activate the audible and visual alarm subroutine. If the values do not exceed the thresholds, the system will continue collecting data. The specific flowchart of the main program is shown in Figure 6.

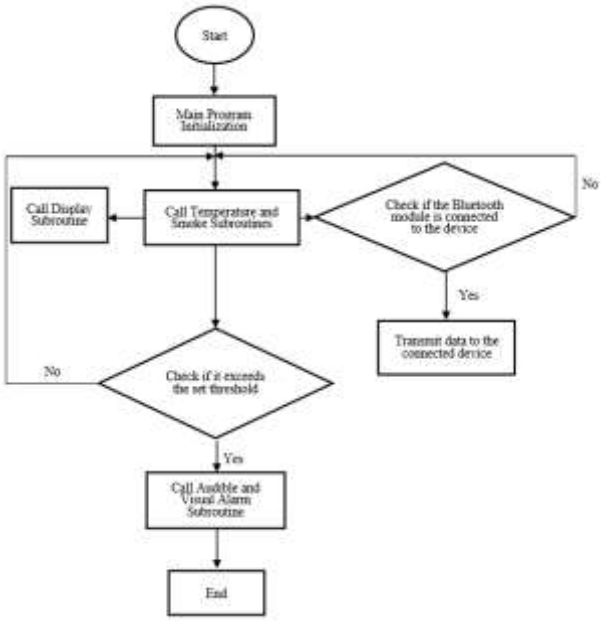


Figure.6 Flowchart of the main system program

International Journal of Science and Engineering Applications
Volume 14-Issue 01, 47 – 51, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1401.1011

4.2 System Core Module Subroutine Design and Flowchart

The system program consists of multiple subroutines, each responsible for completing a specific function. Among them, the most essential subroutines are the temperature sensor subroutine and the smoke sensor subroutine.

The temperature sensor subroutine is an essential component of the entire system program, as outlined below: First, a predefined temperature threshold needs to be set. Then, the temperature sensor continuously monitors the ambient temperature in real-time, comparing the detected temperature data with the preset temperature threshold. If the detected temperature exceeds the threshold, the audible and visual alarm subroutine is triggered to activate the alarm. The flowchart for the temperature sensor subroutine is shown in Figure 7.

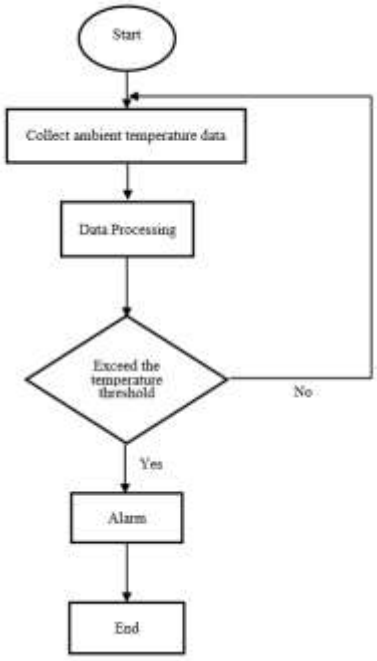


Figure.7 Flowchart of the temperature sensor subroutine

The smoke sensor subroutine is an indispensable part of the system program, as described below: First, a smoke concentration threshold is set. The system then calls the smoke sensor subroutine to control the smoke sensor to collect ambient smoke concentration data, which is compared with the preset smoke concentration threshold. If the detected concentration exceeds the threshold, the system enters the audible and visual alarm subroutine to trigger the alarm. Otherwise, it continues collecting ambient smoke concentration data. The specific flowchart for the smoke sensor subroutine is shown in Figure 8.

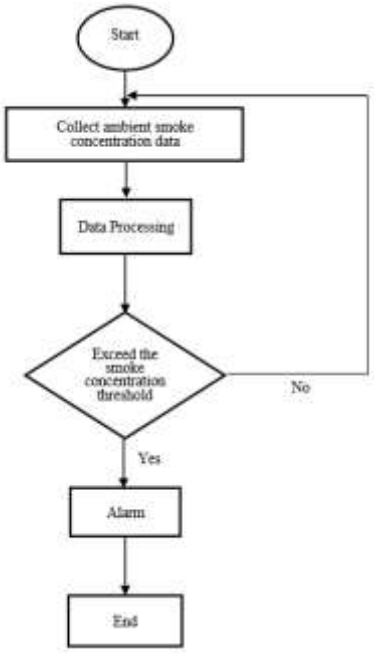


Figure.8 Flowchart of the smoke sensor subroutine

5. SYSTEM DEBUGGING

In the debugging process of alarm system, it mainly involves the verification of temperature monitoring module, smoke concentration monitoring module and Bluetooth communication module. When debugging the temperature monitoring module, touch the temperature sensor to simulate the ambient temperature change and observe whether the system can send an alarm signal in time when the temperature exceeds the set threshold. The debugging of the smoke concentration monitoring module simulates the gas environment of different concentrations by using combustible gases (such as propane, butane, etc.) instead of smoke. Observe whether the system triggers an alarm when the gas concentration exceeds the threshold. When debugging the Bluetooth module, ensure that the alarm can be successfully paired with the mobile device and verify the consistency of the data. By adjusting the threshold and observing the change on the alarm display, confirm the normal operation of the Bluetooth communication module. In addition, data synchronization is performed repeatedly to ensure stable communication. Through repeated debugging, the results show that the system can timely and stably complete various tasks, the system can accurately respond to environmental changes, trigger corresponding alarms, and maintain data consistency in Bluetooth communication.

6. CONCLUSION

This thesis aims to design and implement a temperature and smoke alarm system centered around the STC89C52 microcontroller. The thesis comprehensively presents the entire process from theoretical exploration, system design to final implementation. The system effectively integrates temperature and smoke concentration monitoring, significantly enhancing the timeliness and accuracy of fire hazard warnings. Through

an in-depth analysis of existing fire warning technologies and leveraging the functional advantages of the STC89C52 microcontroller, the proposed design demonstrates efficiency, cost-effectiveness, and stability in both hardware selection and software design.

After rigorous testing and optimization, the designed temperature and smoke alarm system has shown excellent environmental adaptability and reliability. It can effectively provide early warnings in the event of a fire, offering valuable time for evacuation and firefighting efforts. However, the system still has certain limitations. For example, the Bluetooth module's data transmission distance limits the range of remote monitoring. To address this issue, the adoption of a Wi-Fi module could be considered, given its longer transmission range and higher data transfer rate.

7. REFERENCES

[1] Bu Yaqun, Guo Junmei, Liu Haiying. Design and Implementation of Intelligent Fire Alarm System Based on 51 Microcontroller [J]. Journal of Qilu University of Technology, 2021, 35(04): 53-58.

[2] Hao H B, Dai F Z, Wen H K, et al. Research on the Smart Home Design based on Single-chip Microcomputer [C]. Proceedings of the 2020 International Conference on Artificial Life and Robotics (Icarob2020), Beppu, Japan. 2020: 13-16.

[3] Yancheng C, Liheng W. Design and Implementation of Fire Monitoring and Warning User Terminal System [C]. 2022 International Conference on Artificial Intelligence and Computer Information Technology (AICIT). IEEE, 2022: 1-5.

[4] Wang X, Yun H, Cui P, et al. Design of thermometer based on STM32 and Bluetooth [J]. Journal of Computational Methods in Sciences and Engineering, 2021, 21(5): 1417-1432.

[5] Guo W, Feng X. Remote multipoint temperature detection based on single chip microcomputer system design [C]. Third International Conference on Machine Learning and Computer Application (ICMLCA 2022). SPIE, 2023, 12636: 976-980.

International Journal of Science and Engineering Applications
Volume 14-Issue 01, 52 – 57, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1401.1012

Research and Application of AES Algorithm in Symmetric Encryption

Zhonghao Zheng
School of Electronic Information and Electrical Engineering
Yangtze University
Jingzhou, China

Abstract: With the rapid development of information technology, data security issues are becoming more and more prominent, and symmetric encryption, as a common encryption method, has received widespread attention. And among the symmetric encryption algorithms, AES is favored for its high security, efficiency and reliability. This paper firstly introduces the basic principles and technical details of AES algorithm and discusses its specific operation in the process of data encryption and decryption. AES algorithm adopts the substitution-replacement network structure and encrypts the data through multiple rounds of iterative operations, and it has three kinds of key lengths of 128-bit, 192-bit, and 256-bit, which can be selected according to the security requirements, and it has a higher level of security. Then this paper discusses the AES algorithm mainly has four cryptanalysis methods, namely, brute-force cracking, timing attack, linear analysis and differential analysis, as well as the encryption and decryption experiments of the AES algorithm, and the encryption and decryption experiments of the text information using OpenSSL, to evaluate the performance of the encryption and decryption process, and to provide the relevant performance indexes and analysis results. Finally, this paper discusses the application of AES algorithm in various fields. In the financial field, AES is used to encrypt financial transaction data and customer authentication information. In the field of Internet of Things (IoT), AES protects the communication security between IoT devices. In military applications, AES protects the security of military communications and data. These application cases fully demonstrate the importance and wide application of AES in various fields. In summary, this thesis systematically introduces the research and application of AES algorithm in symmetric encryption based on AES algorithm, deeply analyzes the principle, technical details, and applications in various fields of AES algorithm, and puts forward the suggestions for the future research and application. AES algorithm plays an important role in information security protection, and it is of great significance to protect the security and privacy of data.

Keywords: AES Algorithm; Symmetric Encryption; Encryption and Decryption; Cryptanalysis Methods; Application Fields

1. INTRODUCTION

With the rapid development of information technology, the problem of information security has become increasingly prominent. As an important means of information security, symmetric encryption plays a key role in protecting data confidentiality. Limitations of Traditional Encryption Algorithms Early symmetric encryption algorithms, such as DES, have certain limitations in terms of security and efficiency. Therefore, more advanced algorithms need to be researched to meet the growing security needs. Improvement in computing power, with the increasing processing power of computers, attackers are also able to break traditional encryption algorithms more easily. As a result, there is a need to develop stronger and more secure encryption algorithms such as AES. Requirements of regulations and standards Many industries and organizations have developed regulations and standards related to information security that require the use of strong encryption algorithms to protect sensitive information. Together, these factors have led to research on AES algorithms to meet the needs of modern information security.

From 2001 to 2005, domestic scholars began to study AES algorithms and published a series of papers in the field of cryptography, covering the theoretical basis, analysis and improvement of AES algorithms, etc[1]. From 2006 to 2010, with the increase of information security requirements, domestic attention began to focus on the practical application and performance optimization of AES algorithms. Some research focused on hardware implementation, acceleration algorithm and security analysis of AES algorithm, etc[2]. From 2011 to 2015, domestic scholars made some breakthroughs in AES algorithm research and proposed some new encryption schemes and optimization strategies to improve the performance and security of AES algorithm in practical applications[3]. From 2016 to 2020, with the big data and Internet of Things (IoT) technology's rapid development, domestic scholars began to explore the application of AES algorithms in these emerging fields, such as IoT security and cloud computing security. At the same time, the research on AES algorithm in the field of mobile communication and network security is also gradually deepened[4]. From 2021 to 2024 (as of now), AES algorithm is still one of the focuses of cryptography research in China.

From 2000 to 2010, the AES algorithm, as the national standard of the United States, attracted extensive attention from the international cryptography research community. Foreign scholars mainly focused on the security analysis, attack model and countermeasure strategy of AES algorithm, etc[5]. From 2011 to 2015, with the rise of emerging technologies such as cloud computing, Internet of Things and mobile communication, foreign scholars began to study the application and optimization of AES algorithm in these fields. At the same time, new attack methods such as side channel attack and quantum computing attack of AES algorithm are studied[6]. From 2016 to 2020, foreign scholars show a diversified trend in the research direction of AES algorithm. On the one hand,

International Journal of Science and Engineering Applications
Volume 14-Issue 01, 52 – 57, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1401.1012

the hardware implementation, performance optimization and security enhancement of the AES algorithm continue to be explored; on the other hand, the cross-application of cryptography and the exploration of emerging fields become hot spots of research[7]. From 2021 to 2024 (as of now), the AES algorithm remains one of the important topics in international cryptography research. Foreign scholars continue to focus on the security, performance and applicability of the AES algorithm, and apply it to various emerging fields, such as blockchain and artificial intelligence security. In summary, domestic and foreign research on AES algorithm began with its birth and continues to this day. In the past decades, AES algorithms have made great progress in theoretical research, security analysis, performance optimization and practical application, and have made important contributions to the development and progress of the information security field[8].

With the increasing network security threats, the security requirements for data encryption algorithms are also increasing. AES algorithm, as a widely recognized symmetric encryption algorithm, has received more attention and research. With the development of hardware technology, especially the emergence of hardware gas pedals and specialized chips for encryption processing, it has become a hot topic to study how to implement efficient AES encryption algorithms on hardware to improve the encryption speed and energy-efficiency ratio. The AES algorithm is widely used in many standards and protocols, such as TLS and IPsec. Therefore, research on AES algorithms also involves collaboration with standardization organizations and updating and improving the standards. With the development of emerging technologies, such as quantum computing and artificial intelligence, new challenges are posed to the security of traditional encryption algorithms. Therefore, it has become an important direction to study how to improve the AES algorithm's resistance to quantum attacks and its security in the AI environment. The increasing demand for data security in finance, medical care, e-commerce and other fields has driven the research and improvement of AES algorithms in practical applications to meet the changing security needs. Taken together, the background of foreign research on AES algorithms in the field of symmetric encryption is mainly influenced by various factors such as the improvement of security needs, hardware technology development, standardization and normalization, emerging technology challenges and practical application needs. The current direction of AES algorithm research: (1) research on new types of ciphers; (2) research on the principles and guidelines of comprehensive assessment of cryptographic security; (3) research on the implementation of ciphers including software optimization hardware implementation and special chips, etc.; and (4) research on the analysis of the AES and its application.

2. THEORY AND METHODS

2.1 Symmetric encryption fundamentals

Symmetric encryption is an encryption technique used in the field of information security, and its basic concept involves the principle of using the same key for both encryption and decryption processes. In symmetric encryption, the sender uses a key to encrypt the message to form a cipher text and the receiver can decrypt the message to plain text only by using the same key. The key is the same for both the encryption and decryption process and hence it is called symmetric key. The basic principle of symmetric encryption is to encrypt the message using the key so that unauthorized users cannot understand the content of the encrypted message. Only the sender and receiver who know the key can perform the encryption and decryption operations correctly. This design of symmetric encryption makes the data protected from unauthorized access and theft during transmission.

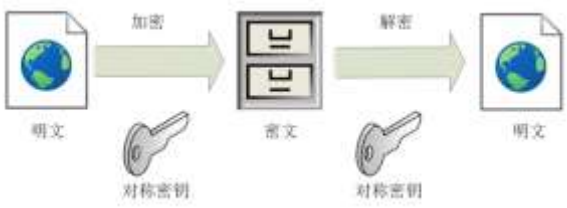


Figure 1 Basic principle diagram of symmetric encryption

The basic principle diagram of symmetric encryption is shown in Figure 1, where a plaintext message is encrypted with a symmetric key to form a ciphertext, and the ciphertext is then decrypted with the same key to form a plaintext message. One of the advantages of symmetric encryption is that it is very fast because the encryption and decryption processes use the same key and do not require overly complex mathematical operations. This makes symmetric encryption efficient in scenarios such as network communication and data transfer. However, symmetric encryption also has some drawbacks, not the least of which is the key management problem. Because the key needs to be shared between the sender and the receiver, the security and management of the key becomes one of the challenges faced by symmetric encryption. To solve the key management problem, symmetric encryption is usually used in combination with other techniques, such as asymmetric encryption to ensure security when transferring keys. Asymmetric encryption uses a pair of keys, public and private, where the public key is used for encryption and the private key is used for decryption. In this way, the sender can encrypt the symmetric key using the receiver's public key while the receiver decrypts the symmetric key using his private key. In conclusion, symmetric encryption is an important encryption technique which has an important role in the field of information security. Even though there are some challenges, such as key management and security issues, symmetric encryption is still one of the effective means to protect data security after combining with other encryption techniques and improving the key management strategy.

2.2 Principles of the AES algorithm

AES is a symmetric-key encryption algorithm, one of the most used encryption algorithms today, and is widely used in the fields of data protection and secure communications. AES is a symmetric-key encryption standard defined by the National Institute of Standards and Technology in 2001 as a replacement for the DES algorithm. The AES algorithm employs the concept of packet ciphers, which divides the data into fixed length chunks and uses the same key for encryption and decryption operations. The AES algorithm adopts the concept of group cipher, dividing data into blocks of fixed length and using the same key for encryption and decryption operations. The operation process is that the plaintext data first undergoes an initial round of processing, including byte substitution, row shifting, column obfuscation, and round key addition, etc. After the initialization round, the data block is divided into multiple columns and then processed through multiple rounds of the round function. Each round of operation includes byte substitution, row shifting, column obfuscation and round key addition, etc. After multiple rounds of wheel function operation, the last round does not include column obfuscation,

International Journal of Science and Engineering Applications
Volume 14-Issue 01, 52 – 57, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1401.1012

but directly performs operations such as byte substitution, row shifting and round key addition, etc., and the encrypted block of data is obtained after the last round of processing and is called the ciphertext.

The features of the AES algorithm are equally numerous. The AES algorithm has won the favor of a wide range of users for its excellent security, efficient performance, flexible key length and wide range of applications. First, the algorithm's high level of security is remarkable, as it can effectively resist all kinds of known cryptographic attack techniques, such as differential analysis and linear analysis, etc., to ensure the confidentiality of the data in the process of transmission and storage. Second, the AES algorithm demonstrates excellent execution efficiency in both hardware and software implementations, enabling both encryption and decryption operations to be performed quickly and in a variety of computing environments. In addition, the AES algorithm provides a variety of key length options, including 128-bit, 192-bit and 256-bit, which allows users to flexibly adjust the key length according to the actual security needs, thus improving the security of the algorithm. At the same time, the increase in key length also makes it significantly more difficult to crack, providing strong support for data security protection. It is worth mentioning that the design structure of AES algorithm is simple and clear, easy to understand and implement. This feature makes the AES algorithm ideal for a variety of secure communication and data protection scenarios, such as network communication, file transfer, database encryption, electronic payment, and virtual private networks. In addition, the AES algorithm is also the basis of many security protocols and standards, such as the SSL/TLS protocol and the IPsec protocol, which provide a solid foundation for building a secure and reliable communication environment. In conclusion, the AES algorithm occupies an important position in the field of data security by virtue of its multifaceted advantages and has become one of the widely used encryption technologies. Both individual users and enterprise organizations can safely adopt AES algorithm to protect their data security.

Advanced Encryption Standard (AES), as a symmetric encryption algorithm, plays a crucial role in securing sensitive data. Whether it is the encrypted transmission of emails, the secure storage of files, or the confidentiality of network communications, AES has demonstrated excellent performance and stability. The basic structure of AES mainly consists of key expansion, wheel function, inverse wheel function, etc.

The basic structure of the AES algorithm is shown in Fig. 2, the first key expansion, according to the key provided by the user, to generate a series of round key, used in the subsequent round of the function of the round key addition operation, the plaintext and the first round of the round key for the bitwise dissimilarity operation, and then after multiple rounds of encryption, each round to go through the byte substitution, row shifting, column obfuscation, the round key addition, and the final round of the final round of the step does not include column obfuscation, and ultimately the final encrypted The final encrypted state is the ciphertext. The decryption process also generates a series of round keys, the same as the encryption process, the initial round of the ciphertext and the last round of the round key for bitwise dissimilarity operation, the decryption process and the encryption process is the opposite, but the order is reversed, i.e., the round key is applied in the reverse order of the order of the final round of the final round of the encryption process with the final round of the same decryption process is the plain text of the state of the decryption.

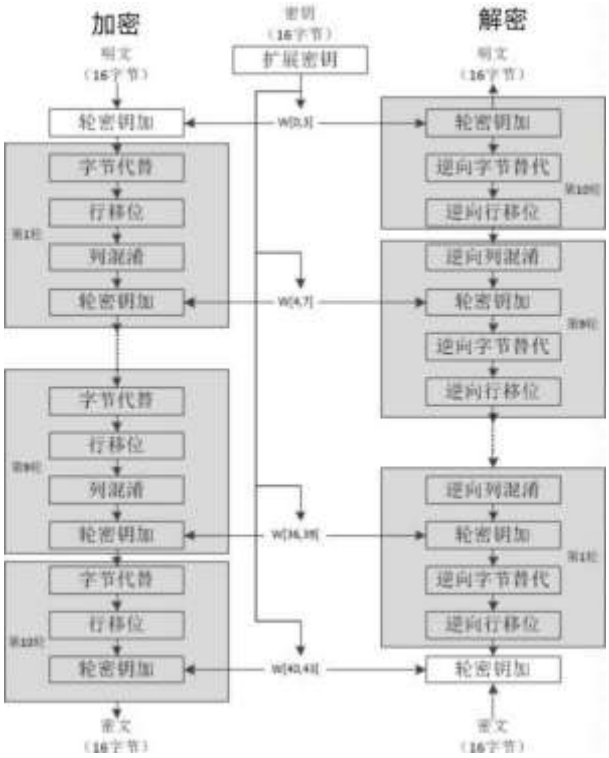


Fig. 2 Basic structure of AES

2.3 Number Theoretic Foundations of the AES Algorithm

AES is a symmetric-key encryption algorithm whose design is based on several key concepts in number theory, especially finite-field arithmetic. The encryption process of AES relies on the structure of Substitution-Permutation Network (SPN)[9], which performs encryption and decryption operations through multiple iterative rounds. The number theoretic foundation of the AES algorithm focuses on addition, multiplication, and polynomial operations over the finite field GF (2^8), which ensure the algorithm's security, complexity, and resistance to attacks.

The AES algorithm uses the finite field GF (2^8) (i.e., modulo 256 operations) for most of the operations in the encryption and decryption process. The finite field GF (2^8) is a field containing 256 elements, where each element is an 8-bit binary number. Addition and multiplication within this domain are performed by binary addition and modulo operations, while multiplication is usually implemented by multiplying polynomials and is performed using an irreducible polynomial for modulo reduction. Specifically, the addition operation employed in AES is the different-or (XOR) operation, which is an addition over the finite domain GF (2), while the multiplication operation is based on the computation of multiplicative inverses, which are used with affine transformations to enhance the nonlinearity of the algorithm. These finite domain operations not only enhance the complexity of the encryption but also make it more difficult for an attacker to crack the key.

Key expansion, a central aspect of the AES algorithm, plays a key role in expanding the short input key into a longer sequence of keys used to encrypt the round function. The algorithm is carefully constructed to ensure that a series of complex and

secure round keys are generated from the input key, which are an integral part of each encryption round. According to the AES standard, the key expansion process determines the number of final round keys strictly based on the length of the input key. Specifically, if the input is a 128-bit key, 10 round keys are generated; if the key length is 192 bits, 12 round keys are generated; and for a 256-bit key, 14 round keys are generated. This design ensures that keys of different lengths are handled appropriately to meet diverse security requirements. Under the action of the key scheduling algorithm, the input keys undergo a series of complex mixing, transforming and rearranging operations to gradually generate the required round keys for each round. These operations are designed to enhance the randomness and complexity of the key, making the generated round key sequence highly secure and unpredictable. As the algorithm advances, a new round key is generated for each round and is added to the round key sequence in an orderly manner. This process is repeated until all round keys are generated. Eventually, we get a complete sequence of round keys, which will play a vital role in the subsequent encryption and decryption processes. The key expansion process ensures the security and efficiency of the AES algorithm and increases the difficulty of searching the key space by generating a complex and highly randomized sequence of round keys, thus improving the security of encryption.

The Round Function is a key part of the AES algorithm that is called during both encryption and decryption. The Round Function consists of four main steps. The byte substitution step replaces each byte with another byte using a fixed Substitution Box, where the substitution rules are fixed and irreversible, which increases the security of the algorithm. Row Shift In this step, each row of the AES state matrix is cyclically shifted left according to a specific rule. Specifically, the first row is kept unchanged to maintain the stability of the data; the second row is shifted to the left by one byte, which realizes the initial exchange of data within the matrix; the third row is shifted to the left by two bytes, which further increases the degree of data obfuscation; and the fourth row is shifted to the left by three bytes to ensure that the bytes of all the rows can be sufficiently obfuscated in the state matrix. Column Obfuscation This step performs an obfuscation operation on each column of the AES state matrix, which is achieved by multiplying it with a fixed matrix. The column obfuscation operation increases the nonlinearity of the algorithm and enhances the anti-analytical performance of AES. Wheel Keys Plus In each round of encryption of the AES algorithm, wheel keys play a crucial role. They are carefully generated by the key expansion algorithm and are designed to increase the randomness and complexity of the encryption process. When the wheel keys are subjected to a bitwise dissimilarity operation with the current state matrix, it not only realizes the obfuscation of the data in the state matrix but also ensures that each round of encryption operation possesses a unique transformation characteristic. The introduction of this step significantly enhances the obfuscation of the AES encryption algorithm such that the output of the same plaintext encrypted with the same key will be different in each round. This increased variability greatly improves the security of the algorithm and makes it more difficult to crack the AES algorithm.

The inverse wheel function phase of AES plays a crucial role in the decryption process, as it can restore the ciphertext to the original plaintext without any errors. In contrast to the wheel function phase of encryption, the inverse wheel function phase performs operations in the opposite order, ensuring that the encryption and decryption processes are complementary. One of the key steps in the inverse wheel function phase is inverse byte substitution, which corresponds to the byte substitution step in the encryption process. In inverse byte substitution, each byte of the ciphertext is carefully processed and each byte is accurately replaced with its corresponding value according to the mapping rules of the inverse S-box. This inverse substitution operation is like a key in the decryption process, which gradually restores the original appearance of the plaintext data and lays the foundation for the subsequent decryption steps. Immediately followed by the reverse row shift step, which is the reverse of the encryption process of the row shift operation. In the retrograde shift, each line is shifted in a reverse cycle according to specific rules, and as the number of lines increases, the amount of shift gradually decreases. This operation precisely adjusts the order of the bytes in each row, restoring them to their pre-encryption state. The execution of the inverse row shifting step not only makes the data in the ciphertext be effectively organized but also provides a guarantee for the final decryption result. In the inverse column obfuscation step, the inverse column obfuscation operation plays a crucial role by precisely performing an inverse linear transformation on each column. This operation effectively reverses the effect of column obfuscation in the encryption process by applying inverse matrix multiplication. By inverting the column obfuscation, the column data that was originally disrupted during the encryption process is recovered and re-presented as it was before the column obfuscation, paving the way for the subsequent decryption step. Immediately after that, the reverse wheel key addition step further advances the decryption process. In this step, the reverse-round key addition operation maintains a high degree of similarity with the round key addition operation in the encryption process. By performing a bitwise different-or operation between the current round's reverse-round key and the decryption state, the reverse-round key addition operation gradually restores the original state of the plaintext data. This operation not only restores the obfuscation level of the data but also ensures the accuracy of the final decryption result. It is important to note that the order of operations in the reverse wheel function stage is completely opposite to that of the wheel function stage in the encryption process. This reverse operation design concept allows the AES algorithm to gradually restore the original plaintext data during the decryption process. Through the well-designed steps of inverse row shift, inverse byte substitution, inverse column obfuscation and inverse wheel key addition, the inverse wheel function stage ensures the integrity and security of the data, providing a solid guarantee for the safe transmission and storage of data.

3. TEXT ENCRYPTION AND VIDEO ENCRYPTION WITH AES ALGORITHM USING OPENSSL

3.1 Experimental platforms and tools

OpenSSL is an open-source cryptography toolkit that provides implementations of various cryptographic functions and protocols, including the SSL/TLS protocol, encryption algorithms, digital certificate management, etc. OpenSSL implements the SSL and TLS protocols, which are used to provide security and privacy protection for network communications. The SSL/TLS protocol provides encryption, authentication, and integrity protection, and is commonly used to protect web applications, email transmissions, etc. OpenSSL supports a variety of encryption algorithms, including symmetric encryption algorithms, asymmetric encryption algorithms, and hash functions. These algorithms can be used to encrypt data, generate digital signatures, calculate message digests, etc. OpenSSL supports the generation, issuance,

verification and management of digital certificates. Digital certificates play an important role in network security and are used for authentication and secure communication. OpenSSL provides a range of command line tools for performing various cryptographic operations. OpenSSL includes interfaces and tools for generating secure random numbers, which are essential for cryptographic operations and key generation. Secure random numbers are used in cryptographic operations for the generation of initialization vectors, random number seeds, keys, etc. OpenSSL runs on a wide variety of operating systems, including Linux, Unix, Windows, etc., enabling developers to use the same cryptographic tools and functionality on different platforms. Overall, OpenSSL is a powerful, flexible and widely used cryptographic toolkit that provides developers with a rich set of cryptographic functions and protocol implementations that help build secure network communications and applications. The experimental environment is configured as follows: The operating system is Windows 10 / Ubuntu 20.04, the OpenSSL version is OpenSSL 1.1.1, and the experimental hardware is an Intel i7 processor with 16GB RAM.

3.2 Experimental data set

In order to verify the encryption effect of the AES algorithm on different data types, two types of files are selected for encryption in this experiment, for text files, a customized text file in .txt format is selected as the experimental data, and for video files, a 5-minute-long video file in .mp4 format is selected, with a resolution of 1920x1080. The purpose of selecting these data is to test the encryption and decryption performance of the AES algorithm when dealing with different sizes and types of files. The purpose of choosing these data is to test the performance of AES algorithm in encrypting and decrypting files of different sizes and types. These data were chosen to test the performance of the AES algorithm in encrypting and decrypting files of different sizes and types. These data have been chosen to test the encryption and decryption performance of the AES algorithm on files of different sizes and types. The performance of the conventional DES algorithm is also compared with the performance of the conventional DES algorithm in processing these files.

3.3 Encryption and decryption process

The core operation of the AES algorithm is based on symmetric key encryption, i.e., the same key is used for encryption and decryption. In this experiment, the AES algorithm with 128-bit key length is used to perform encryption and decryption operations on text and video data. In the text file encryption stage, the specific operation steps first select a text file (input.txt). Then select the key and choose a key of 128-bit (16-byte) length (e.g., “yourpassword”). This key will be used for round key generation during encryption and decryption. The encryption operation is performed using the AES-128 algorithm. Enter the OpenSSL enc -aes-128-cbc -a -in your.txt -out encryted.txt -K yourpassword -iv initialization vector command to encrypt in cbc mode and automatically generate the encrypted.txt file to be used for storing the encrypted file, -K is the passphrase, which consists of 32 hexadecimal digits, and -iv is the initialization vector, can be the same as the password, or you can set it yourself, the length should not be too short, -a is the base64 encoding. available, try the font named Computer Modern Roman. On a Macintosh, use the font named Times. Right margins should be justified, not ragged.

To decrypt an encrypted file, enter the command OpenSSL enc -aes-128-cbc -a -in encrypted.txt -out decode.txt -K yourpassword -iv initialize vectors -d to decrypt the file, automatically generating a decode file to store the decrypted file, -K and -iv must be the same as in encrypted, and -d for the decryption command. The decrypted file is the same as the original file

Video and text encryption have the same instruction structure when using the AES algorithm, because AES is a symmetric encryption algorithm, regardless of whether the encrypted data is text, video, or any other type of data, as long as the same encryption mode is used (e.g., AES-128-CBC), the structure of the instructions to encrypt and decrypt are the same, but they have different data formats and processing performance

3.4 Performance Evaluation

After the encryption and decryption operations were completed, a series of performance evaluations were performed, including the following: measuring the time required from the start of encryption to the completion of encryption. Measure the time needed from the start of decryption to the completion of decryption. Record the change in file size after encryption to evaluate the impact of the encryption algorithm on the file size. Ensure file consistency between encrypted and decrypted files by calculating the hash value of the encrypted file.

3.5 Results

We recorded the time of AES and DES in encrypting and decrypting different files (text files and video files) and the results are shown in Table 1 below.

Table 1. Performance Comparison of AES Algorithm and DES Algorithm

File type	AES encryption time	AES decryption time	DES encryption time	DES decryption time
text file	0.02s	0.01s	0.05s	0.04s
Video files	12.5s	11.8s	20.3s	19.2s

The AES algorithm outperforms DES in handling both data types, along with a slight change in file size before and after encryption. As both AES and DES algorithms padded the file, the file size increased slightly. Specifically, the size of the file after AES encryption increases slightly (e.g., by 1-2 KB) compared to the original file, while the change in file size after encryption is basically the same for DES. We verify the integrity of the encrypted data by hash value comparison. All the AES and DES encrypted files have the same hash value as the original file after decryption, proving that both perform well in terms of confidentiality and data integrity.

4. CONCLUSIONS

In this study, we explore the application of AES algorithm in text and video encryption, focusing on analyzing the experimental process and effect of encryption and decryption through OpenSSL tool. The experimental results show that the AES algorithm, as an efficient and secure symmetric encryption algorithm, can effectively protect the privacy of sensitive data, especially in the encryption process of large-scale data, such as text and video, showing excellent performance. Through the encryption experiments on text and video files, we find that AES has high security and stability in the encryption and decryption process. Especially for the AES-128 mode, the key length is sufficient to meet the daily encryption requirements, and at the same time, the parallel

processing capability of the AES algorithm makes it possible to realize efficient encryption and decryption processes under the condition of hardware acceleration. The encrypted ciphertext, despite the increase in size, the AES algorithm successfully protects the original data content and prevents the risk of data leakage through reasonable key management and encryption mode. Compared with traditional encryption algorithms (e.g., DES, 3DES), the AES algorithm shows obvious advantages in several aspects. First, AES supports longer key lengths (128-bit, 192-bit, 256-bit), which provides stronger resistance to cracking compared to DES's 56-bit key. Second, AES is more computationally efficient during encryption and decryption, and is able to process large-scale data more quickly, especially in the encryption of videos and other large file types, where AES performs more efficiently compared to traditional encryption algorithms. However, despite its outstanding performance in terms of encryption efficiency and security, we still need to pay attention to its actual performance in different environments, such as embedded devices and low-power devices. As emerging technologies such as quantum computing continue to evolve, traditional symmetric encryption algorithms may face new challenges, and future research may need to explore how AES algorithms can enhance their resistance to quantum computing attacks and how their performance can be further improved by optimizing hardware and algorithms. Overall, the AES algorithm, as a core tool for modern information security, has been widely used in multiple fields, such as text encryption and video encryption, and has demonstrated strong technical advantages. In the future, we will continue to explore the performance of AES algorithm in more complex application scenarios in order to further improve its security and efficiency.

5. ACKNOWLEDGEMENT

Many people have given me great help and support during the completion of my thesis. I would like to express my heartfelt thanks to them with immense gratitude. Finally, I would like to thank everyone who has worked hard for this thesis and the authors of all the references, whose research results have provided important theoretical support for my study. It is by standing on the shoulders of the giants before me that I have been able to make gains in this field. Although I have tried my best to do my best in the process of writing and researching the thesis, there are still inevitably shortcomings, and I earnestly ask all experts and teachers to criticize and correct me. Once again, I would like to thank all those who have supported, cared and helped me!

6. REFERENCES

[1] Zhang, H., Li, W., & Liu, Y. (2004). *A Study on the Development and Application of AES Algorithm in Cryptography*. Journal of Cryptographic Research, 12(2), 45-56.

[2] Wang, J., & Chen, X. (2009). *Optimizing AES Algorithm: Performance Enhancements and Security Analysis*. International Journal of Cryptography and Security, 14(4), 113-130.

[3] Liu, Q., Zhang, Z., & Yang, T. (2014). *Research and Application of AES Algorithm: New Encryption Schemes and Optimization Strategies*. Journal of Applied Cryptography, 20(3), 201-214.

[4] Zhao, F., & Xu, P. (2018). *Exploring the Applications of AES Algorithm in IoT and Cloud Computing Security*. International Journal of Network Security, 16(1), 87-101.

[5] Smith, L., & Roberts, D. (2007). *Security Analysis and Countermeasures of AES Algorithm: Attack Models and Defense Strategies*. Cryptography Review, 18(4), 151-165.

[6] Liu, X., & Zhang, Y. (2013). *Hardware Implementation and Performance Optimization of AES Algorithm*. Journal of Hardware Security, 11(2), 120-134.

[7] Johnson, M., & Smith, A. (2018). *The Role of AES in Emerging Technologies: Quantum Computing and AI Security*. Journal of Cryptographic Systems, 22(3), 210-225.

[8] Patel, R., & Kumar, S. (2022). *Application of AES Algorithm in Blockchain and AI Security*. Journal of Modern Cryptography, 29(1), 135-146.

[9] Daemen, J., & Rijmen, V. (2002). AES proposal: Rijndael. In Advanced Encryption Standard (AES) - Design and Analysis (pp. 195-230). Springer.

International Journal of Science and Engineering Applications
Volume 14-Issue 01, 58 – 60, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1401.1013

Research on Optimization of Teaching Resource Push Based on Adaptive Learning Path

Rongyi He*
Information Work Office
Shandong University (Weihai)
Weihai, Shandong, 264209, China

Xiaoqun Wang
Information Work Office
Shandong University (Weihai)
Weihai, Shandong, 264209, China

Abstract:The research on optimization of teaching resource push based on adaptive learning path explored the implementation method of personalized learning in the intelligent education environment. The study proposed a learning path feature framework with "how to learn" and "what to learn" as the core, and designed a resource recommendation and knowledge point matching model by combining static and dynamic features. In terms of "how to learn", the accuracy of recommendation is improved through the diversification of multimedia resources and learning preference analysis; in terms of "what to learn", dynamic feature analysis supports real-time adjustment of learning content to achieve the optimal matching of knowledge points. In addition, the study also explored the application of digital technology in the lifelong education system to provide flexible and efficient learning solutions for individuals at different learning stages. Finally, this study constructed a teaching resource push model that integrates resource push, path optimization and learning behavior feedback, providing theoretical and technical support for personalized education practice.

Keywords: Teaching resource push, study optimization, adaptive learning path

1. INTRODUCTION

As a learning method, autonomous learning has received extensive attention and application in the field of education in recent years. Its theoretical basis is rooted in human creativity and intrinsic motivation, emphasizing the active role played by individuals in the learning process. Autonomous learning is not only a reflection of learners' subjective initiative, but also an important way to cultivate innovation, problem-solving and independent thinking abilities. The core of the autonomous learning is to promote students' transformation from passive acceptance of knowledge to active acquisition and construction of knowledge. This transformation can not only improve learning efficiency, but also enhance students' sense of control and sense of achievement over the content they have learned, thereby enhancing their learning motivation. In the "Opinions on Comprehensively Deepening Curriculum Reform to Implement the Fundamental Task of Moral Education and Cultivation" issued by the Ministry of Education, the requirements for the development of core literacy of Chinese students are proposed, among which "autonomous development" is clearly listed as an important component. This core literacy emphasizes that autonomous learning ability is one of the key abilities that students should possess, and is an important foundation for adapting to future lifelong learning and social development. In this era of rapid development of informatization and globalization, the continuous changes in society and technology have placed higher and higher demands on talents, and the mastery of a single knowledge can no longer meet the needs of future society. Only students with autonomous learning ability can continuously update and deepen their knowledge structure and adapt to the ever-changing social environment.

The implementation of autonomous learning is not achieved overnight. It requires the joint efforts of the education system, teachers and students. The education system should provide students with a good environment for autonomous learning, including innovative curriculum design, rich learning resources and appropriate technical support. Teachers should provide guidance and support to students in teaching, but should not interfere too much in students' autonomous choices and decisions, and help students develop the ability and habit of self-learning. Students themselves also need to have good self-management skills and be able to make effective learning plans and adjustments based on personal interests, learning goals and social needs. Specifically in practice, there are many forms of the autonomous learning. For example, the popularity of online education platforms provides students with a more flexible way of learning. Students can choose the corresponding course content according to their own learning progress and interests. At the same time, the rise of teaching methods such as problem-oriented learning and project-based learning has further promoted the application of autonomous learning and cultivated students' teamwork ability, innovative thinking and practical ability. In the Figure 1, the key aspects of the autonomous learning are illustrated.

International Journal of Science and Engineering Applications
Volume 14-Issue 01, 58 – 60, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1401.1013

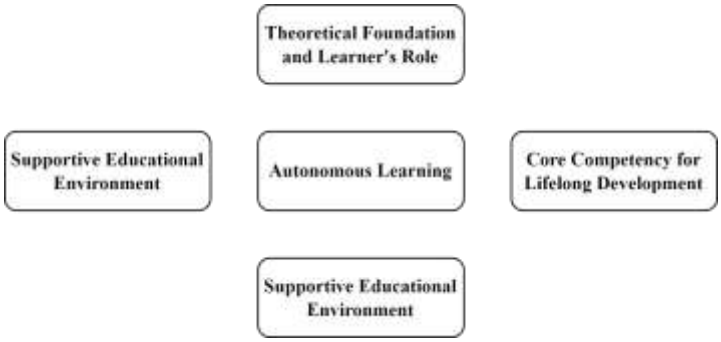


Figure. 1 The Key Aspects of the Autonomous Learning

2. THE PROPOSED METHODOLOGY

2.1 Analysis of the teacher's role in the context of adaptive learning

In the era of smart education, the rapid development of artificial intelligence technology is reshaping the traditional education model, and the role of teachers and the way students learn have undergone profound changes. Teachers are no longer the only controllers of learning content. Students can obtain learning resources through a variety of channels, including traditional offline classrooms, online virtual teachers, and adaptive learning platforms. This transformation not only expands the boundaries of learning, but also greatly improves the personalization and flexibility of learning.

Smart education creates a learning space that combines the real and the virtual, deeply integrating the physical environment with the digital virtual environment. Students can communicate face-to-face with teachers and classmates in the school's physical classroom, or interact with artificial intelligence through knowledge graphs in the virtual classroom. Such a learning space has the following characteristics:

Multi-functional integration: The learning space integrates functions such as resource acquisition, information push, data recording, result evaluation, and analysis and judgment, and builds an intelligent service chain throughout the entire learning process.

Personalized learning support: Through data collection and analysis of students' learning behavior, the system can accurately grasp students' learning preferences and weak links, thereby providing customized learning resources and paths.

Seamless connection: The physical and virtual environments can complement each other to achieve a learning experience anytime, anywhere, and fully meet students' needs for the fragmented learning.

With the support of smart education, students become the leaders of learning. They can choose the learning resources and paths that best suit them, and arrange their learning time and progress independently through the platform. This learning method not only cultivates students' self-discipline and independent thinking ability, but also improves the efficiency and effectiveness of learning.

2.2 Construction of digital lifelong learning education system

As a transformative force driven by digitalization, the power of digital is profoundly affecting the model and connotation of lifelong learning education. From the perspective of technological innovation, human civilization has always evolved along with the development of technology, and breakthroughs in digital technology have brought new possibilities to education, especially in the acquisition, transmission and application of knowledge. The power of digital not only promotes innovation in educational forms, but also reshapes the overall structure of the education system, becoming an important engine for promoting changes in the lifelong learning system.

The digital lifelong education system covers a variety of people, including the first age (children and adolescents), the second age (adults) and the third age (elderly). The differences in cognitive ability, learning goals and learning methods among different groups determine the diversity of educational content and technology applications. For example, the first age emphasizes fun and exploration, the second age focuses more on skill improvement and career development, and the third age focuses on the combination of health, interest and social participation.

Digital technology allows education to present a diversified form, including online courses, virtual laboratories, adaptive learning platforms and personalized learning paths. Behind these types of characteristics is the flexibility and dynamism that digital technology has injected into education. Students can choose the appropriate learning mode according to their personal needs.

The implementation of digitalization runs through all aspects of education, such as analyzing student learning data through artificial intelligence to optimize the distribution of learning content; using knowledge graphs to build the systematic knowledge framework; and recording and verifying learning outcomes through big data and blockchain technology. These means make lifelong learning more efficient, transparent and sustainable.

2.3 Adaptive learning path recommendation

The core features of learning paths can be expanded from two dimensions: "how to learn" and "what to learn", covering the personalized needs of resource recommendation and knowledge point matching respectively.

How to learn: accuracy and diversity of resource recommendation

"How to learn" is mainly based on the characteristics of recommended resource types for learners, including but not limited to learning style, learning preference, media format, interaction method and knowledge granularity. For example, different learners have different acceptance of media such as text, audio, and video. Learning paths need to provide diversified learning resources according to learners'

preferences. At the same time, with the advancement of technology, the addition of emerging media such as virtual reality (VR) and augmented reality (AR) has also brought more possibilities for resource recommendation of learning paths. The selection of these resources should not only meet the current learning needs of learners, but also take into account their future learning potential.

What to learn: knowledge point matching driven by dynamic features

"What to learn" focuses on domain knowledge features, especially the introduction of dynamic features. Dynamic features play a vital role in learning recommendations. They can dynamically adjust recommended content according to learners' real-time learning performance. For example, by analyzing the learner's answer record, learning speed, knowledge mastery and other data, the system can update the difficulty level of the learning path in real time to ensure that the learner is always in the "optimal challenge zone" (Zone of Proximal Development, ZPD). In addition, dynamic features also support the cross-chapter knowledge point correlation analysis to help learners build a more systematic knowledge network.

3. CONCLUSIONS

This study aims to optimize the push of teaching resources in the context of intelligent education and proposes a push model based on adaptive learning paths. By analyzing the characteristics of learning paths, the study emphasizes the importance of two key dimensions: "how to learn" and "what to learn":

In terms of the accuracy and diversity of resource recommendations, by combining the learners' static characteristics (such as media format, interaction mode) and dynamic characteristics (such as learning difficulty, real-time feedback), targeted resource push is achieved, which improves the personalization level of learning experience.

In terms of knowledge point matching and dynamic adjustment of learning paths, through real-time analysis of learning behavior and performance data, the learning path is dynamically optimized, so that learners are always in a suitable challenge range, which promotes the systematic construction of the knowledge system.

In addition, this study emphasizes the key role of digital technology in lifelong education, including the application of multimedia resources, knowledge graphs and blockchain technology in the diversification of educational forms, transparency of results and efficiency of processes. The research results provide innovative ideas for the development of personalized education and promote the improvement of the teaching resource push model in the era of intelligent education. Future research can further combine more complex learning scenarios and large-scale learning data to explore more efficient push mechanisms and evaluation methods.

4. REFERENCES

[1] Champoux, Joseph E. "Animated films as a teaching resource." Journal of management education 25, no. 1 (2001): 79-100.

[2] Miles, Karen Hawley, and Linda Darling-Hammond. "Rethinking the allocation of teaching resources: Some lessons from high-performing schools." Educational Evaluation and Policy Analysis 20, no. 1 (1998): 9-29.

[3] Bizimana, Dr Benjamin, and John Aluko Orodho. "Teaching and learning resource availability and teachers' effective classroom management and content delivery in secondary schools in Huye District, Rwanda." Journal of Education and Practice 5, no. 9 (2014).

[4] Guloba, Madina, James Wokadala, and Lawrence Bategeka. "Does teaching methods and availability of teaching resources influence pupils' performance: evidence from four districts in Uganda." (2010).

[5] Guloba, Madina, James Wokadala, and Lawrence Bategeka. "Does teaching methods and availability of teaching resources influence pupils' performance: evidence from four districts in Uganda." (2010).

[6] Medrea, Nicoleta, and Dana Rus. "Challenges in teaching ESP: Teaching resources and students' needs." Procedia Economics and Finance 3 (2012): 1165-1169.

[7] Edwards, Sarah, and Nick Cooper. "Mind mapping as a teaching resource." The clinical teacher 7, no. 4 (2010): 236-239.

[8] Otieno, Kennedy Omondi. "Teaching/learning resources and academic performance in mathematics in secondary schools in Bondo District of Kenya." Asian social science 6, no. 12 (2010): 126.

[9] Ruthven, Kenneth, Sara Hennessy, and Rosemary Deaney. "Incorporating Internet resources into classroom practice: pedagogical perspectives and strategies of secondary-school subject teachers." Computers & Education 44, no. 1 (2005): 1-34.

[10] Roth, Wolff-Michael, Kenneth Tobin, Cristobal Carambo, and Chris Dalland. "Coteaching: Creating resources for learning and learning to teach chemistry in urban high schools." Journal of Research in Science teaching 41, no. 9 (2004): 882-904.

[11] Pepin, Birgit, Binyan Xu, Luc Trouche, and Chongyang Wang. "Developing a deeper understanding of mathematics teaching expertise: an examination of three Chinese mathematics teachers' resource systems as windows into their work and expertise." Educational studies in Mathematics 94 (2017): 257-274.

[12] Davis, Erin, Dory Cochran, Britt Fagerheim, and Becky Thoms. "Enhancing teaching and learning: Libraries and open educational resources in the classroom." Public Services Quarterly 12, no. 1 (2016): 22-35.

[13] Tschannen-Moran, Megan, and Anita Woolfolk Hoy. "The influence of resources and support on teachers' efficacy beliefs." In annual meeting of the American Educational Research Association, New Orleans, LA. 2002.

[14] Liang, Xi, and Jiesen Yin. "Recommendation algorithm for equilibrium of teaching resources in physical education network based on trust relationship." Journal of Internet Technology 23, no. 1 (2022): 133-141.

International Journal of Science and Engineering Applications
Volume 14-Issue 01, 61 – 64, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1401.1014

Research on the Construction of Oracle Bone Inscriptions Knowledge Graph and its Application in Historical Language Analysis

Dandan Song*
Tianjin Normal University
Tianjin, 300387, China

Changhao Su
Tianjin Normal University
Tianjin, 300387, China

Abstract:As a precious cultural heritage of Chinese civilization, the study of oracle bone inscriptions is of great significance to the fields of philology, history and archaeology. This paper focuses on the construction of oracle bone inscription knowledge graph and its application in historical language analysis, focusing on the text organization, grammatical structure analysis and semantic information mining of oracle bone inscriptions. By adopting natural language processing and machine learning methods, this paper realizes the automatic annotation and deep semantic relationship mining of oracle bone inscriptions. At the same time, the construction of cross-modal knowledge graphs combined with multimodal data provides systematic technical support and multi-dimensional research perspectives for oracle bone inscription research. The research results show that the oracle bone inscription knowledge graph plays a significant role in improving the efficiency of oracle bone interpretation, exploring the social relationship network of the Shang Dynasty and promoting the digital protection of traditional culture.

Keywords: Oracle bone inscriptions; knowledge graph; historical language analysis

1. INTRODUCTION

Oracle bone inscriptions are a treasure of ancient Chinese culture and an important symbol of the development of human civilization. As one of the earliest Chinese characters, oracle bone inscriptions, with their unique artistic form and historical value, occupy an irreplaceable position in the fields of philology, archaeology and art. From its discovery to its research, oracle bone inscriptions are not only regarded as a record of history, but also a cultural symbol that transcends time and space, showing the ancient people's profound understanding and unique expression of nature, society and the universe.

From the perspective of literary art, the shape structure of oracle bone inscriptions is diverse and full of wisdom, covering pictographic, ideographic, phono-semantic and loan forms. These forms not only carry the symbolic meaning of the original pictures, but also develop into a highly abstract symbol system. The pictographic features of the oracle bone script are particularly prominent, and its character shapes often directly imitate the forms of nature. For example, the character "日" (sun) looks like the round sun, and the character "月" (moon) looks like the crescent moon. This intuitive depiction of natural forms gives oracle bone inscriptions a strong pictorial flavor, making them a visual representation of the interaction between ancient peoples and nature. In terms of font structure, the square and round changes in oracle bone script show a unique beauty. Some glyphs have strong and powerful lines and obvious symmetry; while others are flexible and interesting. For example, the character “水” has smooth lines, which seem to show the shape of flowing water; while the character “木” uses an upright shape to show the growth characteristics of trees. This font design not only makes the text expressive, but also provides the source and inspiration for later calligraphy art.

Oracle bone inscriptions are not only a carrier of historical records, but also an important part of Chinese traditional culture. It has passed down the thoughts, beliefs and lifestyles of our ancestors in a symbolic way, becoming an important link between the past and the present. The artistic features displayed by oracle bone inscriptions not only provide rich aesthetic inspiration for future generations, but also inspire modern art and design. Many artists have drawn elements from oracle bone inscriptions and incorporated them into modern calligraphy, font design, and visual art, giving this ancient writing system new vitality.

In addition, the study of oracle bone inscriptions is also of great significance to the development of anthropology, linguistics and philology. It is key material for exploring the origin of Chinese characters and provides an indispensable basis for understanding the formation of the Chinese writing system. By studying oracle bone inscriptions, we can gain a deeper understanding of many aspects of ancient society, including politics, economy, religion, and culture.

Figure. 1 The Example of Oracle Bone Inscriptions

2. THE PROPOSED METHODOLOGY

2.1 The Oracle Knowledge Graph Construction

In recent years, with the rapid development of information technology and artificial intelligence, the research on oracle bone inscription information processing has gradually entered

a stage of in-depth and systematic development, which has not only promoted the digital protection and dissemination of oracle bone inscriptions, a precious cultural heritage, but also provided new research ideas and tool support for related fields. In the study of oracle bone inscriptions, researchers have gradually explored a variety of technical means, combining multidisciplinary methods such as the computer science, linguistics, information visualization and data mining, and have achieved remarkable results, especially in the storage, analysis and application of oracle bone inscription data.

Technical achievements in oracle bone information processing

Construction of oracle bone character database

The establishment of oracle bone character database is the basic work of oracle bone informatization research. By collecting and organizing oracle bone rubbings, literature and interpretation results, researchers have created a standardized digital character database. These character databases not only contain oracle bone characters, interpretations and phonetic information, but also combine contextual data such as time, geography, and sacrifices, providing a solid foundation for multi-dimensional research on oracle bone characters. For example, tools such as the "Oracle Bone Online Dictionary" facilitate researchers to quickly retrieve and analyze oracle bone characters.

Computer-assisted oracle bone splicing

Splicing is an important part of oracle bone research, which is to piece together and restore broken oracle bone pieces. Traditional splicing methods rely on the experience and intuition of experts, which is time-consuming and labor-intensive. Through image recognition and machine learning technology, researchers have developed a computer-assisted splicing tool that can analyze the texture, text position and carving direction of oracle bone pieces through algorithms to achieve automatic matching and splicing, greatly improving research efficiency.

Oracle bone script corpus annotation and editing

In terms of corpus annotation, researchers introduced natural language processing (NLP) technology to systematically annotate the oracle bone script corpus with syntactic structure, semantic role, and contextual information. These annotated data not only help the study of oracle bone script interpretation, but also provide high-quality training data for machine translation and knowledge graph construction.

Oracle bone script machine translation

Oracle bone script machine translation is one of the hot research directions in recent years. By training deep learning models, researchers try to achieve automatic translation from oracle bone script to modern Chinese. This technology is still in the exploratory stage, but has achieved initial results in specific fields (such as common sacrificial terms and royal records).

The construction of oracle bone knowledge graph still faces many challenges, especially in expressing deep semantic relationships.

Mining of deep semantic relationships

Oracle bone knowledge not only includes direct information about glyphs and interpretations, but also contains complex social, religious and political relationships. For example, the lineage of the Shang kings, the geographical locations of various local countries, and the division of responsibilities of different Zhenren. These deep semantic relationships cannot be fully expressed by a simple "entity-attribute-relationship" model. Researchers need to introduce more complex semantic network modeling methods, combining historical background and domain knowledge to achieve higher-level knowledge expression.

Technical support for cross-domain integration

The construction of oracle bone knowledge graph cannot be separated from the theoretical and methodological support of multiple disciplines. For example, metrological citation analysis and co-occurrence analysis can help reveal the vocabulary associations and text contexts in oracle bone inscriptions; information visualization technology can visualize complex knowledge networks. In particular, based on scientific knowledge graphs (such as the MKD method), researchers can more clearly show the core structure, development process and interdisciplinary frontiers of oracle bone research.

Handling of semantic uncertainty

Since some of the contents of oracle bone interpretation are still controversial, its semantic expression has a certain degree of uncertainty. How to deal with these uncertainties in the knowledge graphs, maintaining the scientific nature of the data while allowing for reasonable interpretation space, is one of the current research difficulties. Viable solutions include using probabilistic models or the fuzzy logic techniques to support multiple possible semantic interpretations.

2.2 The Cross-modal Knowledge Graph

Cross-modal knowledge graphs use computer technology to study oracle bone inscriptions, especially the in-depth analysis of grammar, syntax and semantic information of oracle bone inscriptions, which is an important task in the intersection of modern science and technology and traditional culture. In this process, the organization and digitization of oracle bone inscriptions is the starting point of the research, laying the foundation for the subsequent construction and analysis of language models.

The compilation of oracle bone scripts not only covers the interpreted texts of the original oracle bone rubbings, but also includes a large amount of extended materials surrounding oracle bone research, such as historical documents, archaeological records, research monographs, academic reviews, and related educational materials and digital resources. The completeness and accuracy of these data directly determine the depth and breadth of subsequent research.

Integration of interpretation and literature

The compilation of oracle bone inscriptions requires the standardization of the original character forms and interpretations, while conducting in-depth analysis based on the background information in the oracle bone documents. This not only includes the interpretation of individual words, but also involves the restoration of the sentence structure and text context of the oracle bone inscriptions.

Fusion of multimodal data

In addition to the characters themselves, oracle bone script research also requires the integration of multimodal information such as the shape of the oracle bones, carved textures, and excavation sites. These data are constructed

through image processing and database, and form a multi-dimensional research system together with text interpretation.

Digitalization and Visualization

By converting the text data of oracle bone inscriptions into digital format and displaying it through information visualization technology, we can more intuitively understand the relationship network, language structure and semantic features in oracle bone inscriptions. For example, the graphical knowledge network displays the sacrificial activities, geographical distribution and social structure of the Shang Dynasty, providing researchers with a new analytical perspective.

3. CONCLUSION

Based on the needs of oracle bone inscription research, this paper explores the methods and applications of constructing oracle bone inscription knowledge graphs. First, by systematically organizing and standardizing oracle bone inscriptions, a digital corpus with characters, interpretations and contextual information as the core is constructed, which provides a solid foundation for the construction of knowledge graphs. Secondly, this paper combines natural language processing technology with deep learning models to realize grammatical structure analysis, entity recognition and semantic relationship mining, and constructs a multi-level oracle bone inscription knowledge graph. At the same time, by introducing multi-modal data fusion technology, the morphology, texture and excavation information of oracle bone inscriptions are integrated to construct a cross-modal knowledge graph, providing a multi-dimensional analysis framework for oracle bone inscription research.

The study found that knowledge graph technology not only improves the efficiency of oracle bone inscription interpretation and splicing, but also can intuitively display the complex relationship network of Shang Dynasty society, including clan genealogy, geographical distribution and religious activities. In addition, the model based on uncertainty processing proposed in this paper also shows good adaptability in dealing with interpretation ambiguity and semantic ambiguity. In the future, with the further development of information technology, the construction and application of oracle bone knowledge graph will inject new vitality into the digital protection and academic research of traditional culture, and help the inheritance and development of traditional culture in the modern context.

4. REFERENCES

[1] Wang, Mengru, Yu Cai, Li Gao, Ruichen Feng, Qingju Jiao, Xiaolin Ma, and Yu Jia. "Study on the evolution of Chinese characters based on few-shot learning: From oracle bone inscriptions to regular script." Plos one 17, no. 8 (2022): e0272974.

[2] Liu, Kexin, Xiaohong Wu, Zhiyu Guo, Sixun Yuan, Xingfang Ding, Dongpo Fu, and Yan Pan. "Radiocarbon dating of oracle bones of late Shang period in ancient China." Radiocarbon 63, no. 1 (2021): 155-175.

[3] Fu, Xuanming, Zhengfeng Yang, Zhenbing Zeng, Yidan Zhang, and Qianting Zhou. "Improvement of oracle bone inscription recognition accuracy: A deep learning perspective." ISPRS International Journal of Geo-Information 11, no. 1 (2022): 45.

[4] Chang, Xiang, Fei Chao, Changjing Shang, and Qiang Shen. "Sundial-gan: A cascade generative adversarial networks framework for deciphering oracle bone inscriptions." In Proceedings of the 30th ACM International Conference on Multimedia, pp. 1195-1203. 2022.

[5] Li, Bang, Qianwen Dai, Feng Gao, Weiye Zhu, Qiang Li, and Yongge Liu. "HWOBC-a handwriting oracle bone character recognition database." In Journal of Physics: Conference Series, vol. 1651, no. 1, p. 012050. IOP Publishing, 2020.

[6] Hung, Shyh-Shiun, Hen-Hsen Huang, and Hsin-Hsi Chen. "A complete shift-reduce Chinese discourse parser with robust dynamic oracle." In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 133-138. 2020.

[7] Gao, Junheng, and Xun Liang. "Distinguishing oracle variants based on the isomorphism and symmetry invariances of oracle-bone inscriptions." IEEE Access 8 (2020): 152258-152275.

[8] Xuzheng, Lu, Cai Hengjin, and Lin Li. "Recognition of oracle radical based on the capsule network." CAAI transactions on intelligent systems 15, no. 2 (2020): 243-254.

[9] Li, Jing, Qiu-Feng Wang, Kaizhu Huang, Xi Yang, Rui Zhang, and John Y. Goulermas. "Towards better long-tailed oracle character recognition with adversarial data augmentation." Pattern Recognition 140 (2023): 109534.

[10] Wang, Mei, Weihong Deng, and Cheng-Lin Liu. "Unsupervised structure-texture separation network for oracle character recognition." IEEE Transactions on Image Processing 31 (2022): 3137-3150.

[11] Yue, Xuebin, Hengyi Li, Yoshiyuki Fujikawa, and Lin Meng. "Dynamic dataset augmentation for deep learning-based oracle bone inscriptions recognition." ACM Journal on Computing and Cultural Heritage 15, no. 4 (2022): 1-20.

[12] Han, Wenhui, Xinlin Ren, Hangyu Lin, Yanwei Fu, and Xiangyang Xue. "Self-supervised learning of Orc-Bert augmentator for recognizing few-shot oracle characters." In Proceedings of the Asian Conference on Computer Vision. 2020.

[13] Guo, Ziyi, Zihan Zhou, Bingshuai Liu, Longquan Li, Qingju Jiao, Chenxi Huang, and Jianwei Zhang. "An Improved Neural Network Model Based on Inception-v3 for Oracle Bone Inscription Character Recognition." Scientific Programming 2022, no. 1 (2022): 7490363.

[14] Liu, Mengting, Guoying Liu, Yongge Liu, and Qingju Jiao. "Oracle bone inscriptions recognition based on deep convolutional neural network." Journal of image and graphics 8, no. 4 (2020): 114-119.

[15] Li, Jing, Qiu-Feng Wang, Rui Zhang, and Kaizhu Huang. "Mix-up augmentation for oracle character recognition with imbalanced data distribution." In Document Analysis and Recognition–ICDAR 2021: 16th International Conference, Lausanne, Switzerland, September 5–10, 2021, Proceedings, Part I 16, pp. 237-251. Springer International Publishing, 2021.

[16] Fujikawa, Yoshiyuki, Hengyi Li, Xuebin Yue, C. V. Aravinda, G. Amar Prabhu, and Lin Meng. "Recognition of oracle bone inscriptions by using two deep learning models." International Journal of Digital Humanities 5, no. 2 (2023): 65-79.

[17] Han, Simon J., Piers Kelly, James Winters, and Charles Kemp. "Simplification is not dominant in the evolution of Chinese characters." Open Mind 6 (2022): 264-279.

International Journal of Science and Engineering Applications
Volume 14-Issue 01, 65 - 72, 2025, ISSN:- 2319 - 7560
DOI:10.7753/IJSEA1401.1015

A Novel Deep Learning Model for Fault Detection in Power Transformers

Nagaraju Brahmanapally
Student, Department of CS
University of West Florida
Pensacola, FL 32514, USA

Drake Fulton
Student, Department of ECE
University of West Florida
Pensacola, FL 32514, USA

Dr. Bhuvana Ramachandran
Professor, Department of ECE
University of West Florida
Pensacola, FL 32514, USA

Abstract: Power transformers are vital in ensuring the reliability of electrical power systems, necessitating accurate fault classification for their efficient operation. This research evaluates a novel Transformer Deep Learning model architecture for fault classification using dissolved gas analysis (DGA) data, leveraging feature engineering and an over-sampling technique to address high-dimensionality and class imbalance challenges. The model demonstrated substantial accuracy improvements across datasets of varying sizes and preprocessing stages, particularly with SMOTE-enhanced data. These findings underscore the effectiveness of Transformer deep learning architectures in advancing the state-of-the-art in fault classification for power transformer systems.

Keywords: Power Transformer Fault Classification, Dissolved Gas Analysis , Transformer Deep Learning Model, Fault Type Prediction, Gas Concentration Analysis, Feature Engineering, Synthetic Minority Over-sampling Technique, Imbalanced Dataset Handling

1. INTRODUCTION

Power transformers are critical components in electrical power systems, responsible for the efficient transmission and distribution of electricity. The reliability of these transformers is paramount to ensuring a stable and uninterrupted power supply. However, transformers are prone to various types of faults, often indicated by the presence of dissolved gases in the transformer oil. The accurate classification of these faults based on gas concentration levels is crucial for timely maintenance and prevention of catastrophic failures. Traditional methods for fault diagnosis, such as the Duval Triangle method and Key Gas Analysis, have limitations in handling the complex, high-dimensional data typically encountered in modern power systems. This has led to the exploration of advanced machine learning and deep learning techniques for more accurate and automated fault classification [1].

In recent years, Transformer models have gained significant attention in various domains, including natural language processing and time-series analysis, due to their ability to capture long-range dependencies and complex relationships within data [2]. The self-attention mechanism, which is central to Transformer architectures, enables these models to weigh the importance of different input features dynamically, making them highly effective in tasks requiring nuanced understanding of input data. In the context of power transformer fault diagnosis, the application of Transformer models is relatively novel. Their potential to handle high-dimensional data and learn intricate patterns makes them promising candidates for improving fault classification accuracy, especially when dealing with data that exhibits significant variability and noise.

Recent research in the field of power transformer fault diagnosis has begun to explore the use of advanced deep learning techniques. For example, Zhi Li et al. [3] proposed a fault diagnosis technique based on Long Short-Term Memory (LSTM) neural networks combined with dissolved gas analysis (DGA). Their study, which analyzed 240 samples, demonstrated that the LSTM model achieved superior diagnostic accuracy compared to traditional neural networks. This underscores the potential of deep learning models to improve fault diagnosis in power transformers. Despite these advancements, there remains a gap in the application of Transformer models specifically for predicting power transformer fault types, suggesting a novel direction for future research [4].

In addition to the challenges posed by high-dimensional data, another critical issue in transformer fault diagnosis is the class imbalance often present in the data. Certain fault types may occur less frequently, leading to a skewed distribution that can bias machine learning models towards the more common classes. To address this issue, Synthetic Minority Over-sampling Technique (SMOTE) has been widely adopted as a data augmentation strategy. SMOTE generates synthetic samples for the minority class by interpolating between existing samples, thereby balancing the class distribution and enabling the model to learn from a more representative dataset [5]. The effectiveness of SMOTE has been demonstrated in various domains, including medical diagnosis, fraud detection, and power systems, where it has been used to enhance the performance of classifiers in imbalanced datasets [6,7]. In the power transformer domain, SMOTE, combined with feature engineering, can significantly improve the model's ability to correctly identify rare but critical fault types.

Several studies have applied Transformer models and SMOTE in different domains with positive outcomes. For instance, Transformer models have been used in the healthcare sector to predict patient outcomes based on time-series data, demonstrating superior accuracy compared to traditional recurrent neural networks (RNNs) [8]. Similarly, SMOTE has been successfully employed in fraud detection tasks to address the issue of imbalanced datasets, leading to more accurate and reliable predictions [6]. Despite these successes, there has been limited exploration of these techniques in the power transformer fault diagnosis domain. This study aims to bridge this gap by evaluating the performance of a Transformer-based model on a dataset of gas concentrations, with a particular focus on the impact of SMOTE and feature engineering on classification accuracy.

International Journal of Science and Engineering Applications
Volume 14-Issue 01, 65 - 72, 2025, ISSN:- 2319 - 7560
DOI:10.7753/IJSEA1401.1015

By leveraging the strengths of Transformer models and addressing the challenges of imbalanced datasets through SMOTE, this research seeks to advance the state-of-the-art in power transformer fault classification. The results presented in this paper not only demonstrate the efficacy of these methods in this domain but also provide insights into their potential application in other critical infrastructure systems where fault diagnosis is essential for maintaining operational reliability.

2. DATASETS

To implement the proposed deep learning model for identifying and classifying faults in transformers, three datasets were collected and processed. The small dataset was obtained from [9], manually entered in a spreadsheet, and saved as a CSV file. This dataset had no missing values and was 100% complete. The medium dataset, collected from [10], contained several missing values for gas concentrations. To ensure data consistency and prevent skewed results, rows with missing values were removed before the dataset was used in the model. This dataset was also saved as a CSV file. Additionally, the large dataset was sourced from [11] and combined with data from [10] to create the most extensive dataset. To address any skewed distributions and optimize the performance of the transformer model, all datasets underwent standardization. This process, which transforms data to have a zero mean and unit standard deviation, enhances the effectiveness of the algorithms.

2.1 Dataset Overview and Preprocessing

2.1.1 Small Dataset

The small dataset comprises 70 samples, each containing six key gas concentration features: hydrogen (H2), methane (CH4), acetylene (C2H2), ethylene (C2H4), ethane (C2H6), and carbon monoxide (CO), all measured in parts per million (ppm). Additionally, the dataset includes a target variable labeled "Fault," which is categorized into four distinct classes: "Thermal," "High Discharge," "Low Discharge," and "No Fault". These fault classes represent the specific fault types to be predicted. The distribution of these fault classes within the dataset is presented in Figure 1.

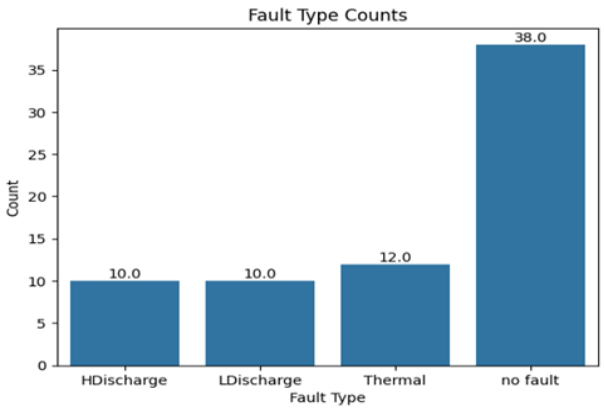


Figure 1: Fault distribution in small dataset

2.1.2 Medium Dataset

The medium dataset initially comprised 151 data points, each containing seven gas concentration features: hydrogen (H2), methane (CH4), acetylene (C2H2), ethylene (C2H4), ethane (C2H6), carbon monoxide (CO), and carbon dioxide (CO2), measured in parts per million (ppm), along with a target variable labeled "Fault". The fault types in this dataset included: 'D1' (Low Energy Discharge), 'D2' (High Energy Discharge), 'None' (No fault), 'HThermal' (High Thermal—thermal faults exceeding 700oC, as determined by equipment inspection), 'LThermal' (Low Thermal—thermal faults below 700oC, as determined by equipment inspection), and 'PD' (Partial Discharge). However, 37 data points had missing values for these gas features. To prevent these missing values from negatively impacting the algorithm's performance and introducing bias, these data points were excluded from the dataset. This likely resulted from unrecorded measurements. After cleaning, the final dataset consisted of 114 complete data points, which were used for subsequent analysis. The distribution of fault types is depicted in Figure 2.

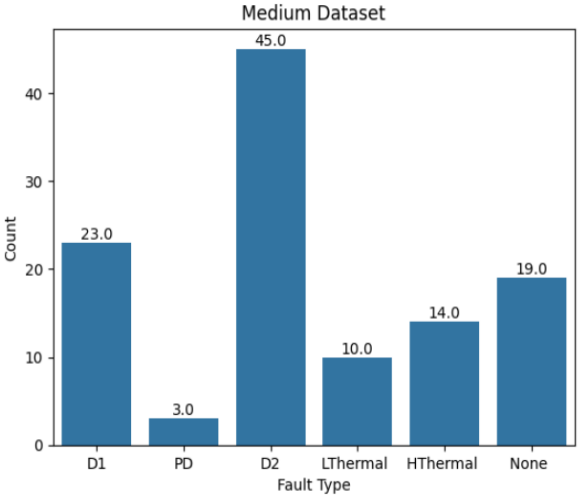


Figure 2: Fault distribution in medium dataset

2.1.3 Large Dataset

The large dataset originally consisted of 231 data points, each containing five gas concentration features: hydrogen (H2), methane (CH4), acetylene (C2H2), ethylene (C2H4), ethane (C2H6), all measured in parts per million (ppm), along with a target variable labeled "Fault." However, 18 data points had missing values for these gas features. To ensure the accuracy of the algorithm and prevent data bias, these incomplete data points were excluded, likely due to unrecorded measurements. After this data cleaning process, the final dataset comprised 213 complete data points, suitable for further analysis. The dataset includes nine distinct fault types, as shown in Table 1. The distribution of these fault types is depicted in Figure 3.

Table 1: Faults and number association

Fault Type	Number
Partial Discharge	0
Spark Discharge	1
Arc Discharge	2
High -temperature Overheating	3
Middle -Temperature	4
Low -Temperature Overheating	5
Low/Middle -Temperature	6
High Energy Discharge	7
No Fault	8

International Journal of Science and Engineering Applications
Volume 14-Issue 01, 65 - 72, 2025, ISSN:- 2319 - 7560
DOI:10.7753/IJSEA1401.1015

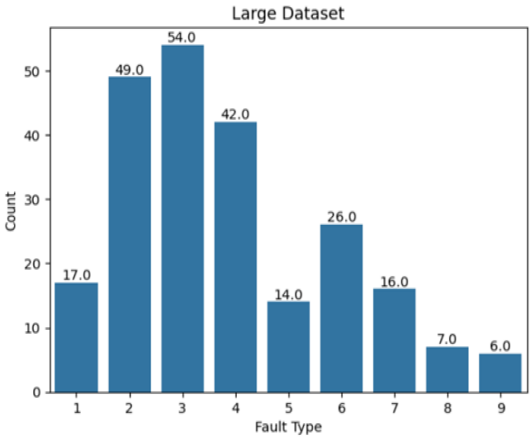


Figure 3: Fault distribution in large dataset

2.2 Feature Engineering: Unveiling Critical Insights

To enhance the performance of the transformer deep learning model, feature engineering was conducted based on a thorough correlation analysis of the gas concentration features within each dataset. Since the gas measurements varied across the datasets, the features generated through the feature engineering process were unique to each case. The correlation matrix heat maps revealed varying degrees of correlation between the features. Features with strong positive correlations and weak negative correlations were used to derive new features through ratio calculations. This strategy was designed to leverage the inherent relationships between gas concentrations while minimizing the influence of features with weaker or opposing trends. Specifically, features with moderate to high positive correlations (greater than 0.7) and weak negative correlations (around -0.1) were selected to create these new ratio-based features. Consistent patterns of correlations were observed across all three datasets.

2.2.1 Small Dataset

In the small dataset, only ratios derived from highly positive correlations were identified and utilized to generate additional features. Six key correlations were selected to create new ratio-based features:

• H2:C2H6 Ratio: The strong positive correlation (corr = 0.92) between Hydrogen(H2) and Ethane (C2H6) suggests a close relationship in their concentrations under fault conditions. This ratio captures and leverages this inherent link.

• H2:CO Ratio: A strong positive correlation (corr = 0.78) exists between Hydrogen (H2) and Carbon (CO) indicating that a higher H2:CO ratio may reflect similar trends in fault-related gas emissions.

• CH4:C2H6 Ratio: The strong positive correlation (corr = 0.87) between Methane (CH4) and Ethane (C2H6) reflect their tendency to vary together, making this ratio a valuable feature for distinguishing fault conditions.

• CH4:CO Ratio: The positive correlation (corr = 0.79) between Methane (CH4) and Carbon (CO) suggest that their concentrations rise and fall together, potentially offering predictive insights through the CH4:CO ration.

• C2H2:C2H4 Ratio: The strong positive correlation (corr = 0.79) between Acetylene (C2H2) and Ethylene (C2H4) highlights their mutual response to fault conditions, making this ratio a meaningful feature for fault classification.

• C2H6:CO Ratio: The strong positive correlation (corr = 0.80) between Ethane (C2H6) and Carbon (CO) further emphasizes the relationship between these gases, allowing this ratio to capture relevant fault-related interactions.

The correlation heatmap for the small dataset is represented in Figure 4.

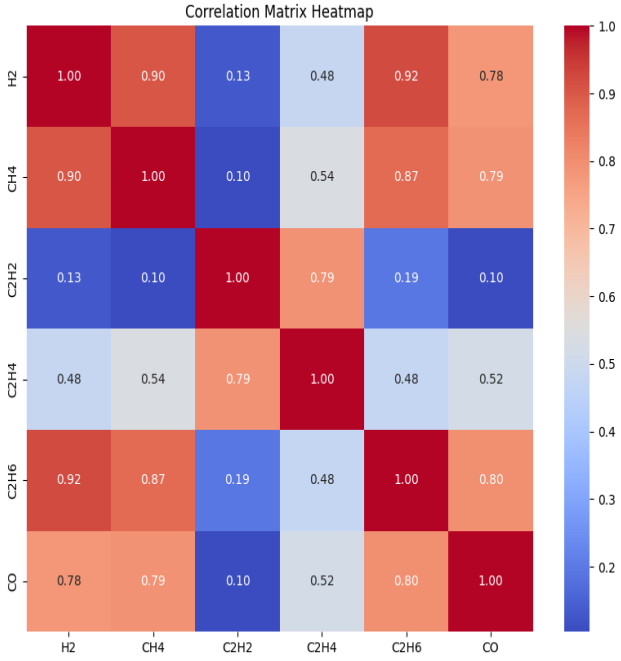


Figure 4: Correlation heat map of the small dataset

2.2.2 Medium Dataset

In the medium dataset, both ratios with high positive correlations and those with weak negative correlations were identified and used to generate additional features. Five key correlations were selected for feature creation:

• H2:CO2 Ratio: This ratio captures the relative concentration of Hydrogen (H2) to Carbon Dioxide (CO2). Although the correlation is weakly negative (corr = -0.06), a higher H2:CO2 ratio may still be indicative of specific fault types.

• CH4:C2H4 Ratio: The strong positive correlation (corr = 0.85) between Methane (CH4) and Ethylene (C2H4) highlights their potential interdependence during fault conditions, making this ratio an informative feature for fault prediction.

• C2H2:CO2 Ratio: Similar to the H2:CO2 ratio, this feature (corr = -0.09) represents the relative concentration of Acetylene (C2H2) to Carbon Dioxide (CO2). Despite the weak negative correlation, this ratio could provide subtle insights into fault characteristics.

• C2H4:C2H6 Ratio: The high positive correlation (corr = 0.76) between Ethylene (C2H4) and Ethane (C2H6) indicates that their concentrations tend to increase or decrease together, providing valuable information for the model through the C2H4: C2H6 ratio.

• CO: CO2 Ratio: The positive correlation (corr = 0.70) between Carbon Monoxide (CO) and CO2 suggests a co-dependent behavior of these gases under transformer fault conditions, making the CO:CO2 ratio a key feature in the dataset.

The correlation heatmap for the medium dataset is illustrated in Figure 5.

International Journal of Science and Engineering Applications
Volume 14-Issue 01, 65 - 72, 2025, ISSN:- 2319 - 7560
DOI:10.7753/IJSEA1401.1015

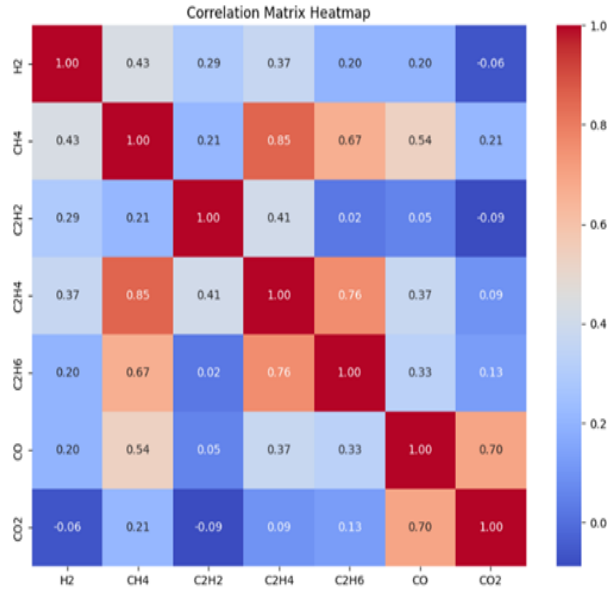


Figure 5: Correlation heat map of the medium dataset

2.2.3 Large Dataset

In the large dataset, only ratios reflecting high positive correlations were identified. Three key correlations were established to create additional features:

• CH4:C2H6 Ratio: This ratio represents the relative concentration of Methane (CH4) to Ethane (C2H6). The strong positive correlation (corr = 0.78) suggests that an increased CH4:C2H6 ratio may indicate similar trends in gas emissions.

• CH4:C2H4 Ratio: A strong positive correlation (corr = 0.87) exists between Methane (CH4) and Ethylene (C2H4), indicating a potential connection in their concentrations during fault conditions. This ratio serves to leverage the inherent relationship within the dataset.

• C2H4:C2H6 Ratio: The very high positive correlation (corr = 0.92) between Ethylene (C2H4) and Ethane (C2H6) signifies that their concentrations tend to rise and fall in tandem under specific fault conditions, providing valuable insights for the model.

The correlation heatmap for the large dataset is presented in Figure 6.

2.3 Data Augmentation and Balancing with SMOTE

2.3.1 Synthetic Minority Oversampling Technique (SMOTE)

The Synthetic Minority Over-Sampling Technique (SMOTE) is a widely used approach for addressing class imbalance in machine learning tasks. Proposed by Chawla et al. [12] in their paper "SMOTE: Synthetic Minority Over-sampling Technique," SMOTE generates synthetic examples of the minority class to balance the distribution of classes in the training dataset. This method aims to enhance the model's ability to learn from underrepresented classes by providing a more even representation of all classes [12].

SMOTE operates by creating synthetic samples in the feature space rather than simply duplicating existing samples. It selects samples that are close in the feature space and generates new samples along the line segments connecting them. This process effectively increases the density of the minority class and improves the model's performance in detecting and classifying underrepresented fault types [12].

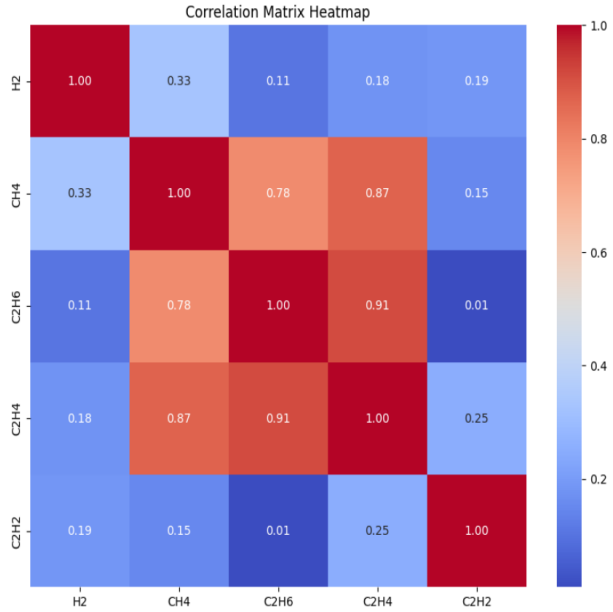


Figure 6: Correlation heat map of the large dataset

The application of SMOTE spans various domains, including medical diagnosis, fraud detection, and industrial equipment fault prediction. In medical diagnostics, SMOTE has been used to improve the detection of rare diseases and anomalies by generating synthetic patient data [13]. In fraud detection, SMOTE helps in identifying fraudulent transactions in imbalanced financial datasets [14]. For industrial applications, such as power transformer fault prediction, SMOTE addresses the challenge of class imbalance by enhancing the model's ability to detect and predict rare but critical fault types [15]. By incorporating SMOTE, predictive models can achieve better performance and more reliable predictions in scenarios where class imbalance is a significant issue.

2.3.2 Data Augmentation with SMOTE

Across all datasets, instances of class imbalance were observed. Such imbalances can lead to inconsistencies during the training of the proposed algorithm, as unbalanced data may result in inaccurate predictions due to the dominance of oversampled classes. Each of the three datasets contained minority classes that could contribute to misclassifications. To address these imbalances, the Synthetic Minority Oversampling Technique (SMOTE) was employed. This technique was applied to all three datasets. Figures 7, 8, and 9 illustrate the distributions of fault types following the application of SMOTE, which equalized the number of samples for each fault type, resulting in a balanced representation across all classes.

International Journal of Science and Engineering Applications
Volume 14-Issue 01, 65 - 72, 2025, ISSN:- 2319 - 7560
DOI:10.7753/IJSEA1401.1015

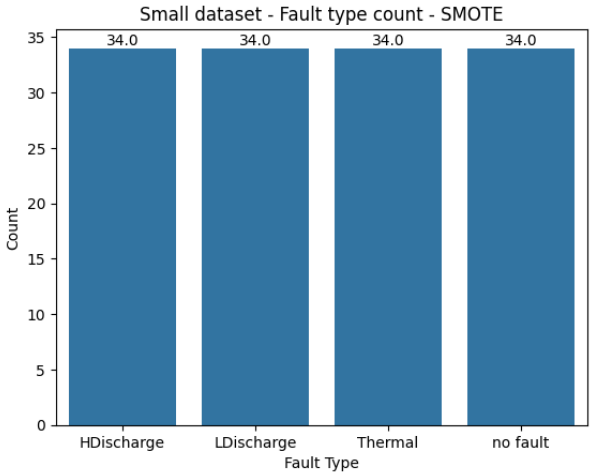


Figure 7: Fault distribution in small dataset after Feature engineering & SMOTE

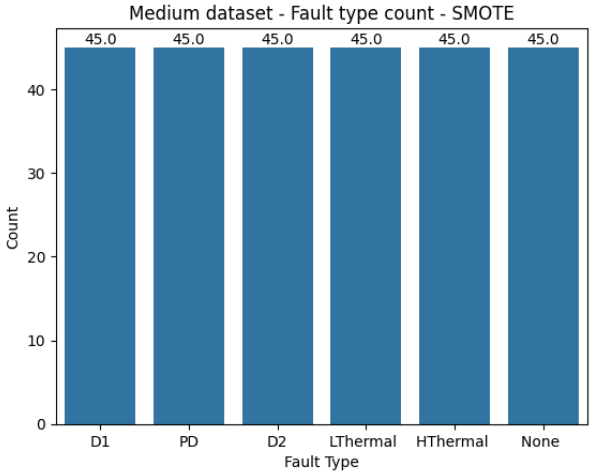


Figure 8: Fault distribution in medium dataset after SMOTE

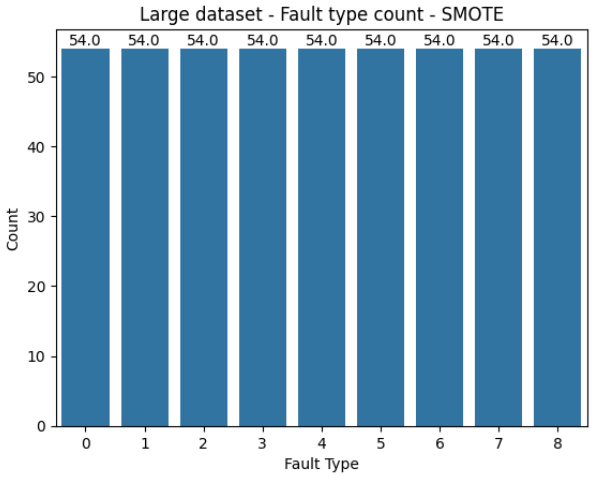


Figure 9: Fault distribution in large dataset after SMOTE

3. DEEP LEARNING TRANSFORMER MODEL

Transformer models represent a significant advancement in the field of deep learning, particularly in handling sequence-to-sequence tasks. Introduced by Vaswani et al. [2] in their seminal paper "Attention Is All You Need," transformers utilize a self-attention mechanism that enables the model to weigh the importance of different elements in the input sequence, irrespective of their position. This mechanism is crucial for capturing long-range dependencies and contextual relationships within data [2].

The architecture of transformers consists of encoder and decoder layers, each equipped with multi-head self-attention and feed-forward neural networks. This design allows transformers to process sequences in parallel, significantly improving efficiency compared to previous sequential models like RNNs and LSTMs [2]. Transformers have achieved state-of-the-art results in various domains, including natural language processing (NLP) and computer vision. In NLP, models like BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer) have set new benchmarks in tasks such as text classification, translation, and summarization [16][17].

In the context of time-series and predictive maintenance, transformers offer promising capabilities. Their ability to handle complex temporal dependencies makes them suitable for analyzing sensor data and predicting faults in industrial systems. Recent research has demonstrated the effectiveness of transformers in fault detection and prediction for various types of equipment, including power transformers [18]. The versatility of transformer model (shown in Figure 10) in capturing intricate patterns and relationships in data positions them as a powerful tool for improving fault prediction accuracy.

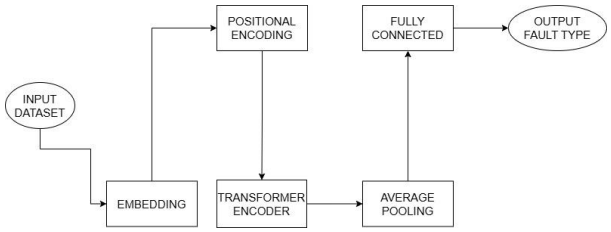


Figure 10: Flowchart of the Deep Learning Transformer model

The Transformer model is designed to process input features representing gas concentrations, which are first subjected to an embedding layer to map the input data into a higher-dimensional space. This process enhances the model's ability to capture the intricate relationships between different gas concentrations. The embedded inputs are then supplemented with positional encoding, which is crucial for retaining the order and structure of the input sequence, despite the inherent lack of sequential information in the input data. Figure 10 how the flowchart of how the deep learning model works.

The core of our model is the Transformer encoder, which leverages multi-head self-attention mechanisms and feedforward neural networks to extract deep, context-aware features from the input data. The use of multiple encoder layers allows the model to learn hierarchical representations of the input, which is essential for accurate fault classification.

To further refine the extracted features, an average pooling operation is applied, reducing the dimensionality and focusing on the most relevant information. The pooled features are then passed through a fully connected layer, which serves as the final classifier, outputting the predicted fault type.

To ensure the robustness of our model, we employed several data preprocessing steps, including the use of Synthetic Minority Over-sampling Technique (SMOTE) to address class imbalance and the removal of missing values to improve data quality. The dataset was then split into training and testing sets, with the training data used to optimize the model's parameters through backpropagation. The model was trained for 200 epochs, with performance monitored through accuracy metrics. The results demonstrate that the Transformer model is capable of achieving significant improvements in fault classification accuracy, surpassing traditional machine learning approaches and bringing us closer to our goal of 80% or more accuracy.

The methodology described above was systematically applied to three different datasets of varying sizes to evaluate the scalability and robustness of the Transformer model. Each dataset was processed through multiple stages, including feature engineering and synthetic data generation using SMOTE, to assess the model's performance across diverse data configurations. The following table 2 summarizes the number of rows and columns for each dataset across the different stages:

Table 2: Dataset sizes

Dataset size	Stage	Dimensions
Small	Original	(70, 7)
	Feature Engineering	(66, 14)
	SMOTE + Feature Engineering	(136, 14)
Medium	Original	(114, 8)
	Feature Engineering	(114, 13)
	SMOTE + Feature Engineering	(270, 13)
Large	Original	(231, 6)
	Feature Engineering	(213, 9)
	SMOTE + Feature Engineering	(465, 9)

In the small dataset, the original data consisted of 70 rows and 7 columns, representing the concentrations of dissolved gases. After applying feature engineering techniques, the dataset was expanded to 66 rows and 14 columns, where additional features were derived to capture more complex patterns. Applying SMOTE to this engineered dataset resulted in a larger dataset of 136 rows and 14 columns, addressing the class imbalance issue and providing a more comprehensive training set for the Transformer model.

The medium-sized dataset began with 114 rows and 8 columns. Feature engineering increased the dimensionality to 13 columns while maintaining the same number of rows. After applying SMOTE, the dataset expanded to 270 rows, further enriching the training data. Similarly, the large dataset, which initially had 231 rows and 6 columns, was transformed through feature engineering into a dataset with 213 rows and 9 columns. SMOTE application resulted in an expanded dataset with 465 rows and 9 columns, providing ample data for model training.

By following this methodology across datasets of varying sizes, we were able to demonstrate the Transformer model's adaptability and consistency in handling different data volumes. The results from these experiments validate the model's potential for generalizing across different datasets, making it a robust tool for fault classification in power transformers. This approach also highlights the importance of data preprocessing and feature engineering in enhancing the performance of deep learning models.

4. IMPLEMENTATION OF TRANSFORMER MODEL FOR FAULT PREDICTION IN TRANSFORMERS

The Transformer model was evaluated on three datasets of varying sizes, each subjected to different stages of data processing: original, feature-engineered, and SMOTE-enhanced feature engineering.

The following algorithm delineates the systematic approach employed for the classification and prediction of fault types in power transformers utilizing a Transformer-based deep learning framework.

4.1 Algorithm:

Step 1: Data Preparation

- Three datasets—categorized as small, medium, and large—were curated, as elaborated in Section 2.1 (Data Overview and Preprocessing).
- These datasets were selected to assess the scalability, robustness, and efficacy of the proposed Transformer model.

Step 2: Data Preprocessing

- The raw datasets were subjected to preprocessing to eliminate missing values (NAs) and to incorporate additional derived features through feature engineering, as detailed in Sections 2.2 and 2.3.
- Feature engineering was undertaken to enhance the representational capacity of the input data and to facilitate the extraction of meaningful patterns for fault classification.

Step 3: Data Partitioning

- The preprocessed datasets were partitioned into training (80%) and testing (20%) subsets to facilitate model training and performance evaluation.
- This stratified division ensures the reliability and generalizability of the proposed methodology.

Step 4: Data Standardization

- Standardization was applied post-class imbalance resolution (via SMOTE) and feature engineering to normalize the datasets.
- Each feature was transformed to have a mean of 0 and a standard deviation of 1. This ensures the mitigation of scale disparity across features, accelerates model convergence, and enhances predictive performance.

Step 5: Model Architecture and Training

The standardized data was fed into a Transformer-based architecture comprising the following sequential components:

i. **Embedding Layer:** Encodes input features into dense vector representations to facilitate model comprehension.

ii. **Positional Encoding:** Introduces positional context into the embeddings, ensuring the model captures the inherent ordering of features.

iii. **Transformer Encoder:** Leverages self-attention mechanisms and feed-forward networks to model intricate dependencies and relationships within high-dimensional data.

iv. **Average Pooling Layer:** Aggregates feature representations to create a compact latent space representation.

v. **Fully Connected Layer:** Maps the latent representation to a high-level feature space for fault classification.

vi. **Output Layer:** Outputs a probability distribution over fault types, with dimensionality corresponding to the number of fault categories.

Step 6: Model Evaluation and Performance Comparison

- The performance of the Transformer-based model was assessed across three dataset variants:
    - Original dataset.
    - Feature-engineered dataset.
    - SMOTE-enhanced dataset.
- Accuracy metrics were computed for each dataset variant, and comparative analysis was conducted to evaluate the impact of feature engineering and data augmentation (via SMOTE) on classification efficacy.

This implementation underscores the viability of Transformer-based deep learning architectures in addressing the challenges of high-dimensional and imbalanced datasets for fault diagnosis in power transformers. The proposed methodology advances the state-of-the-art in fault classification by leveraging feature engineering, SMOTE, and self-attention mechanisms to achieve superior predictive accuracy.

The results, summarized in Table 3, reveal significant variations in accuracy across the different datasets and preprocessing stages, reflecting the impact of data preparation and augmentation on model performance.

For the small dataset, the Transformer model achieved perfect accuracy (100%) on both the original and feature-engineered versions, indicating that the model was able to learn the relationships within the gas concentrations effectively without requiring additional synthetic data. However, when SMOTE was applied to address class imbalance, the accuracy dropped to 71.43%. This decrease suggests that while SMOTE successfully increased the dataset's size, it may have introduced noise or less representative samples that hindered the model's performance.

In the medium dataset, the original dataset yielded a moderate accuracy of 60.87%, which improved slightly to 65.22% after feature engineering. This improvement highlights the benefits of generating additional features to capture more complex relationships in the data. The most significant gain was observed when SMOTE was applied, with the accuracy jumping to 88.89%. This substantial improvement demonstrates the effectiveness of SMOTE in enhancing the training set's representativeness, allowing the Transformer model to generalize better to unseen data.

Table 3: Accuracies after Implementation of Transformer model for Fault Prediction in Transformers

Dataset Size	Stage	%Accuracy
Small	Original	100
	Feature Engineering	100
	SMOTE + Feature Engineering	71.43
Medium	Original	60.87
	Feature Engineering	65.22
	SMOTE + Feature Engineering	88.89
Large	Original	61.7
	Feature Engineering	74.42
	SMOTE + Feature Engineering	89.25

The large dataset exhibited similar trends, with the original dataset yielding an accuracy of 61.7%, which increased to 74.42% after feature engineering. This result underscores the importance of feature engineering in improving model performance, particularly when dealing with larger datasets. The application of SMOTE further boosted the accuracy to 89.25%, highlighting the importance of addressing class imbalance in large datasets. The hyperparameter adjustments made for the SMOTE-enhanced dataset, particularly the reduction in the number of heads and layers, likely contributed to the model's ability to handle the more complex and diverse training data effectively.

The choice of hyperparameters played a crucial role in the performance of the Transformer model across the different datasets and processing stages. For the small and medium datasets, consistent hyperparameters were applied, including a hidden dimension of 64, feed-forward dimension of 128, four attention heads, four layers, and a dropout rate of 0.1. The input dimension and number of classes were adjusted according to the dataset's specific characteristics, with the input dimension ranging from 6 to 13 and the number of classes from 4 to 6. The large dataset required more careful tuning, particularly after applying SMOTE. For the original large dataset, four attention heads and four layers were maintained, but for the feature-engineered and SMOTE-enhanced datasets, the number of heads was reduced to 2, and the number of layers to 3, reflecting the need for a more streamlined architecture to handle the increased data complexity. Across all datasets, a learning rate of 0.001 and 200 training epochs were used, ensuring sufficient training time for convergence without overfitting.

Overall, these results underscore the importance of data preprocessing and augmentation in enhancing the performance of deep learning models. The consistent improvements observed after applying feature engineering

and SMOTE across all dataset sizes validate the robustness and adaptability of the Transformer model in classifying power transformer fault types. The model's performance on the large SMOTE-enhanced dataset indicates its potential for deployment in real-world scenarios where data diversity and class imbalance are common challenges.

5. CONCLUSION

This study demonstrates the effectiveness of using a Transformer-based model for classifying power transformer fault types based on gas concentration levels. By evaluating the model across three datasets of varying sizes and processing stages—original, feature-engineered, and SMOTE-enhanced—it was evident that both feature engineering and data augmentation significantly contributed to improved model accuracy. The model performed exceptionally well on the small dataset, achieving 100% accuracy in the original and feature-engineered stages, though the accuracy dropped to 71.43% after applying SMOTE. The medium and large datasets also showed substantial improvements with the application of SMOTE, with accuracies of 88.89% and 89.25%, respectively, indicating the model's potential to generalize well across diverse and imbalanced data.

The results underscore the importance of comprehensive data preprocessing, careful hyperparameter tuning, and the use of advanced deep learning techniques like Transformers in tackling complex classification tasks in the power systems domain. The hyperparameter adjustments made for the large dataset, particularly in reducing the number of attention heads and layers after applying SMOTE, highlight the necessity of optimizing model architecture to handle increased data complexity effectively.

However, the reliance on synthetic data generated through SMOTE raises concerns about the model's real-world applicability. While SMOTE helps to balance the dataset and improve model performance, it can introduce synthetic patterns that do not entirely represent real-world scenarios. Therefore, future work should focus on gathering more real-time data that captures various gas concentration levels and associated fault types in power transformers. This would enable a more robust evaluation of the Transformer model's performance in practical settings and ensure that the predictions are not overly influenced by synthetic data patterns. Expanding the dataset with real-world measurements will provide a more reliable basis for deploying this model in operational environments, ultimately enhancing its utility in preventing power transformer failures.

6. ACKNOWLEDGMENTS

The authors wish to thank the University of West Florida for providing the funds and facilities to conduct this research.

7. REFERENCES

[1] Duval, M. (2002). A review of faults detectable by gas-in-oil analysis in transformers. _IEEE Electrical Insulation Magazine_, 18(3), 8-17.

[2] Vaswani, A., et al. (2017). Attention is All You Need. In _Proceedings of the 31st International Conference on Neural Information Processing Systems_, 5998–6008.

[3] Li, Z., et al. (2023). Research on the transformer fault diagnosis method based on LSTM artificial neural network and DGA. _2022 International Conference on Intelligent Computing and Machine Learning (2ICML)_, IEEE, 2023.

[4] Zhang, Y., et al. (2022). Fault diagnosis of transformer using artificial intelligence: A review. _Frontiers in Energy Research_, 10, 1006474.

[5] Chawla, N. V., et al. (2002). SMOTE: Synthetic Minority Over-sampling Technique. _Journal of Artificial Intelligence Research_, 16, 321–357.

[6] He, H., & Garcia, E. A. (2009). Learning from Imbalanced Data. _IEEE Transactions on Knowledge and Data Engineering_, 21(9), 1263-1284.

[7] Wang, S., & Yao, X. (2012). Multiclass Imbalance Problems: Analysis and Potential Solutions. _IEEE Transactions on Systems, Man, and Cybernetics_, 42(4), 1119-1130.

[8] Y.T. Yu, M.F. Lau. (2005). A comparison of MC/DC, MUMCUT and several other coverage criteria for logical decisions, Journal of Systems and Software, 2005, in press.

[9] Illias, H.A., Chai, X.R., Abu Bakar, A.H. and Mokhlis, H., 2015. Transformer incipient fault prediction using combined artificial neural network and various particle swarm optimisation techniques. PloS one, 10(6), p.e0129363.

[10] Duval, M. and DePabla, A., 2001. Interpretation of gas-in-oil analysis using new IEC publication 60599 and IEC TC 10 databases. IEEE Electrical Insulation Magazine, 17(2), pp.31-41.

[11] Seifeddine, S., Khmais, B. and Abdelkader, C., 2012, March. Power transformer fault diagnosis based on dissolved gas analysis by artificial neural network. In 2012 first international conference on renewable energies and vehicular technology (pp. 230-236). IEEE.

[12] N. V. Chawla et al., (2002). SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research, vol. 16, pp. 321-357.

[13] G. Lemaître, L. Nogueira, and J. L. D. Carvalho,. (2017). SMOTE: Synthetic Minority Over-sampling Technique. Journal of Machine Learning Research, vol. 18, pp. 1-23.

[14] V. K. Elaparthy, N. R. S. T. S. Suresh, and S. M. Sharmila, (2019). A hybrid SMOTE model for detecting fraudulent transactions in imbalanced datasets. Journal of King Saud University-Computer and Information Sciences.

[15] M. Ahmed, M. Ganaie, and R. Bhat, (2020). An improved hybrid model for transformer fault detection and prediction using SMOTE. Journal of Electrical Engineering & Technology, vol. 15, no. 2, pp. 625-635.

[16] Devlin, J., (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

[17] A. Radford et al., (2018). Improving Language Understanding by Generative Pre-Training. OpenAI.

W. Zhang et al., (2020). Transformer-based approach for fault diagnosis in power systems. IEEE Transactions on Industrial Electronics, vol. 67, no. 8, pp. 6718-6727.

International Journal of Science and Engineering Applications
Volume 14-Issue 01, 73 – 75, 2025, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1401.1016

Application of Artificial Intelligence (AI) in Predicting Mechanisms and Reaction Rates in Chemistry

Tien V. Pham
School of Chemistry and Life Sciences
Hanoi University of Science and Technology
Hanoi City, Vietnam

Abstract: Artificial Intelligence (AI) has emerged as a transformative tool in the field of chemistry, offering unprecedented capabilities in predicting reaction mechanisms and reaction rates. This paper reviews recent advancements in AI methodologies applied to these aspects, focusing on machine learning models, neural networks, and their integration with quantum chemical calculations. The synergy between AI and experimental chemistry is also explored, highlighting its potential to accelerate the discovery of novel reactions and optimize industrial processes.

Keywords: Artificial Intelligence; machine learning; reaction mechanisms; reaction rates; chemistry

1. INTRODUCTION

The rapid advancements in artificial intelligence (AI) have revolutionized various scientific disciplines, with chemistry being no exception. In recent years, AI has emerged as a powerful tool for addressing complex challenges in chemical research, particularly in understanding and predicting reaction mechanisms and rates. Traditional methods for studying chemical reactions, such as quantum mechanical calculations and experimental approaches, are often time-consuming, resource-intensive, and limited in scope. AI offers an innovative alternative by leveraging vast datasets and advanced computational models to analyze, predict, and optimize chemical reactions with remarkable speed and accuracy.[1-3]

This integration of AI into chemistry enables researchers to uncover insights into reaction pathways, transition states, and kinetic parameters that were previously difficult or impossible to determine. From deep learning algorithms that predict reaction outcomes to generative models that design novel reaction pathways, AI is reshaping the way chemists approach problem-solving. Moreover, these advancements have significant implications for industries such as pharmaceuticals, materials science, and green chemistry, where understanding and optimizing reaction rates are critical to innovation and efficiency.[4,5]

This paper explores the application of AI in predicting mechanisms and reaction rates in chemistry, focusing on its methodologies, challenges, and future prospects. By delving into the intersection of AI and chemical research, we aim to highlight the transformative potential of these technologies in accelerating scientific discovery and fostering sustainable development.

2. AI TECHNIQUES IN CHEMISTRY

2.1 Machine Learning Models

Machine learning (ML) models play a critical role in predicting reaction mechanisms and rates by analyzing vast amounts of chemical data. Key techniques include:

Linear Regression and Polynomial Regression: These methods are used for simple reaction systems where relationships between variables are linear or slightly nonlinear. They are particularly useful for initial exploratory analyses of rate constants.

Support Vector Machines (SVM):[6] SVMs are effective for classification tasks, such as determining whether a reaction will proceed under given conditions. They work by identifying hyperplanes in high-dimensional spaces that separate different classes of chemical behaviors.

Random Forests and Gradient Boosting Machines:[7] These ensemble methods excel at capturing complex, nonlinear relationships between molecular descriptors (e.g., atomic charges, bond lengths) and reaction outcomes. Random forests provide interpretability by highlighting the importance of specific descriptors.

Gaussian Process Regression (GPR):[8] GPR is widely used in chemistry for its ability to model uncertainties in predictions. It is particularly useful in active learning scenarios where new experiments are iteratively designed to improve model accuracy.

Kernel Ridge Regression (KRR):[9] KRR is employed for its balance of flexibility and computational efficiency, making it suitable for medium-sized datasets in predicting reaction energies and barriers.

To improve accuracy, these models often rely on curated datasets that include molecular features such as:

Molecular fingerprints[10] (e.g., Extended Connectivity Fingerprints, ECFPs).

Quantum chemical descriptors[11] (e.g., HOMO-LUMO gap, partial charges).

Thermodynamic properties[12] (e.g., enthalpies, entropies).

Data preprocessing steps, including normalization, feature selection, and dimensionality reduction (e.g., via principal component analysis), are critical to enhancing model performance.

2.2 Deep Learning

Deep learning architectures, such as graph neural networks (GNNs)[13] and recurrent neural networks (RNNs),[14] are well-suited for chemistry applications. GNNs model molecules as graphs, where atoms are represented as nodes and chemical

bonds as edges. These networks can predict reaction mechanisms by learning transformations of molecular graphs. Convolutional neural networks (CNNs) are also employed for tasks involving image-based inputs, such as reaction condition optimization through high-throughput experimentation data.

RNNs, particularly in the form of sequence-to-sequence models, have been used to predict reaction outcomes by encoding chemical reaction sequences and learning relationships between reactants and products. Variational autoencoders (VAEs) and generative adversarial networks (GANs) extend these capabilities by enabling the generation of new molecules or reaction pathways.

2.3 Hybrid Approaches

Hybrid approaches combine the strengths of AI with quantum chemical methods. For instance, AI models can predict reaction barriers by interpolating between quantum mechanical calculations, thus reducing computational costs. Quantum chemical data, such as density functional theory (DFT)[15] results, are often used to train AI models, providing accurate predictions of energy profiles and transition states. This integration is particularly valuable in catalysis research, where detailed mechanistic insights are required.

3. APPLICATIONS OF AI IN REACTION MECHANISM PREDICTION

3.1 Mechanistic Pathway Identification

AI models can predict plausible reaction pathways by analyzing the structural and electronic properties of reactants. Tools like Chemprop[16] and ReactionPredictor[17] have demonstrated success in identifying pathways for organic reactions, including pericyclic and photochemical reactions. By leveraging molecular graph representations and advanced machine learning algorithms, these tools can predict how reactants will interact, the intermediates formed, and the products generated.

An example is the use of GNNs for retrosynthetic analysis, where AI predicts the sequence of reactions needed to synthesize a target compound. Such analyses consider not only the thermodynamic feasibility but also the kinetic accessibility of reaction steps, enabling chemists to design efficient synthetic routes.

For complex organic transformations, AI models trained on large reaction databases, such as Reaxys[18] or the USPTO[19] dataset, provide predictions that incorporate solvent effects, temperature, and pressure conditions. This makes them indispensable tools for both academic and industrial research.

3.2 Catalysis and Enzyme Reactions

AI has been instrumental in understanding catalytic mechanisms, both homogeneous and heterogeneous. In homogeneous catalysis, neural networks have been applied to predict the behavior of transition metal complexes, including ligand coordination and activation energy barriers. This aids in the rational design of catalysts with improved efficiency and selectivity.

For heterogeneous catalysis, convolutional neural networks (CNNs)[20] have been used to analyze surface adsorption phenomena, where reactants interact with catalytic surfaces. By integrating AI with computational techniques like density functional theory (DFT), researchers can predict reaction pathways on catalytic surfaces with high accuracy, optimizing processes like ammonia synthesis or CO_2 reduction.

Enzymatic reactions have also benefited from AI, particularly in protein engineering. Machine learning models predict how mutations in enzyme structures will affect their catalytic activity, enabling the design of enzymes with tailored functionalities. For example, AI has been used to design enzymes for biofuel production by optimizing the degradation of lignocellulosic biomass.

3.3 Photochemical-electrochemical Reaction

AI techniques are being applied to predict mechanisms in photochemical and electrochemical reactions, where the involvement of excited states or electron transfer processes adds complexity. Machine learning models trained on high-throughput experimental and theoretical data can predict key properties like redox potentials, excited-state lifetimes, and charge transfer rates. This accelerates the discovery of materials for solar energy conversion, such as organic photovoltaics and photocatalysts.

3.4 Multistep Reaction Networks

In complex reaction networks, such as those encountered in metabolic pathways or polymerization processes, AI models excel at identifying dominant pathways and rate-limiting steps. By integrating kinetic modeling with machine learning, researchers can simulate the dynamic behavior of reaction networks under various conditions, providing insights into system-level properties and emergent behaviors.

4. AI IN REACTION RATE PREDICTION

4.1 Kinetic Modeling

AI has significantly advanced the modeling of reaction kinetics by leveraging extensive datasets of experimental rate constants. Machine learning models, such as random forests and neural networks, are trained to predict rate constants based on molecular descriptors and reaction conditions. These models outperform traditional methods by capturing nonlinear relationships and identifying subtle dependencies.

Deep learning techniques, like graph neural networks (GNNs),[13] enable the direct use of molecular structures as input, learning intricate details about how molecular features influence reaction rates. Active learning strategies further enhance these models by iteratively improving predictions through targeted experimental data acquisition.

4.2 Temperature and Pressure Dependence

Predicting reaction rates across varying temperatures and pressures is a challenging task that AI excels at. Traditional approaches, such as the Arrhenius equation, provide approximations but often fail for complex systems. AI models, trained on high-dimensional datasets that include temperature and pressure variations, offer more precise predictions.

For instance, Gaussian process regression (GPR) and neural networks have been used to map the effects of environmental factors on rate constants of a chemical reaction. These models are particularly effective in catalysis and combustion chemistry, where extreme conditions play a critical role in reaction dynamics. Furthermore, AI can account for secondary effects, such as solvent interactions and reaction intermediates, to refine predictions. Especially, it can figure out rate constants for reaction channels without passing via any tigh transition states. For example, reactions between free hydrocarbon radicals (C_3H_3 and CH_3).

4.3 Predicting Reaction Orders and Rate Laws

Machine learning algorithms can infer reaction orders and rate laws directly from experimental data, bypassing the need for

manual derivation. By analyzing time-series data of reactant concentrations, AI models can determine how changes in concentration influence the overall reaction rate, providing insights into the underlying mechanism.

4.4 High-Throughput Screening

AI-driven high-throughput screening has enabled rapid exploration of reaction conditions to optimize rates. By integrating AI with robotic automation, researchers can test thousands of reaction conditions in a fraction of the time required by conventional methods. This approach has been particularly impactful in pharmaceutical and materials chemistry, where reaction rate optimization is critical for process efficiency.

5. CHALLENGES AND LIMITATIONS

5.1 Data Quality and Availability

The reliability of AI models depends on the quality and diversity of training datasets. Incomplete or biased datasets can lead to inaccurate predictions.

5.2 Interpretability

AI models, particularly deep learning networks, often function as black boxes, making it difficult to interpret the underlying chemical principles driving predictions.

5.3 Generalization

Many AI models struggle to generalize beyond their training data, particularly for reactions involving exotic or novel substrates.

6. FUTURE DIRECTIONS

6.1 Integration with Experiment

AI can be integrated with high-throughput experimentation to generate real-time data for model training and validation, enabling iterative improvements in predictive accuracy.

6.2 Explainable AI

Developing interpretable AI models will enhance their acceptance in the chemistry community and facilitate the discovery of novel mechanistic insights.

6.3 Open-Access Databases

Establishing comprehensive, open-access reaction databases will address data scarcity and improve the robustness of AI models.

7. CONCLUSION

AI represents a paradigm shift in the prediction of reaction mechanisms and rates. By reducing computational costs and accelerating discovery, it holds the potential to revolutionize chemistry. However, addressing challenges related to data quality, interpretability, and generalization will be essential for realizing its full potential.

8. REFERENCES

[1] Ananikov, V. P. 2024. Top 20 Influential AI-Based Technologies in Chemistry. Artificial Intelligence Chemistry, 100075.

[2] Tiwari, P. C.; Pal, R.; Chaudhary, M. J.; Nath, R. 2023. Artificial intelligence revolutionizing drug development: Exploring opportunities and challenges. Drug Development Research, 84(8), 1652-1663.

[3] Brown, N.; Ertl, P.; Lewis, R.; Luksch, T.; Reker, D.; Schneider, N. (2020). Artificial intelligence in chemistry and drug design. Journal of Computer-Aided Molecular Design, 34, 709-715.

[4] Daher, W.; Diab, H.; Rayan, A. 2023. Artificial intelligence generative tools and conceptual knowledge in problem solving in chemistry. Information, 14(7), 409.

[5] Wood, C. 2006. The development of creative problem solving in chemistry. Chemistry Education Research and Practice, 7(2), 96-113.

[6] Mammone, A.; Turchi, M.; Cristianini, N. 2009. Support vector machines. Wiley Interdisciplinary Reviews: Computational Statistics, 1(3), 283-289.

[7] Callens, A.; Morichon, D.; Abadie, S.; Delpey, M.; Liquet, B. 2020. Using Random forest and Gradient boosting trees to improve wave forecast at a specific location. Applied Ocean Research, 104, 102339.

[8] Wang, B.; Chen, T. 2015. Gaussian process regression with multiple response variables. Chemometrics and Intelligent Laboratory Systems, 142, 159-165.

[9] Maalouf, M.; Homouz, D. 2014. Kernel ridge regression using truncated newton method. Knowledge-Based Systems, 71, 339-344.

[10] Cereto-Massagué, A.; Ojeda, M. J.; Valls, C.; Mulero, M.; Garcia-Vallvé, S.; Pujadas, G. 2015. Molecular fingerprint similarity search in virtual screening. Methods, 71, 58-63.

[11] Wang, L.; Ding, J.; Pan, L.; Cao, D.; Jiang, H.; Ding, X. 2021. Quantum chemical descriptors in quantitative structure–activity relationship models and their applications. Chemometrics and Intelligent Laboratory Systems, 217, 104384.

[12] Van Speybroeck, V.; Gani, R.; Meier, R. J. 2010. The calculation of thermodynamic properties of molecules. Chemical Society Reviews, 39(5), 1764-1779.

[13] Corso, G.; Stark, H.; Jegelka, S.; Jaakkola, T.; Barzilay, R. 2024. Graph neural networks. Nature Reviews Methods Primers, 4(1), 17.

[14] Grossberg, S. 2013. Recurrent neural networks. Scholarpedia, 8(2), 1888.

[15] Orio, M.; Pantazis, D. A.; Neese, F. 2009. Density functional theory. Photosynthesis research, 102, 443-453.

[16] Heid, E.; Greenman, K. P.; Chung, Y.; Li, S. C.; Graff, D. E.; Vermeire, F. H.; McGill, C. J. 2023. Chemprop: a machine learning package for chemical property prediction. Journal of Chemical Information and Modeling, 64(1), 9-17.

[17] Kayala, M. A.; Baldi, P. 2012. ReactionPredictor: prediction of complex chemical reactions at the mechanistic level using machine learning. Journal of chemical information and modeling, 52(10), 2526-2540.

[18] Lawson, A. J., Swienty-Busch, J., Géoui, T., & Evans, D. (2014). The making of reaxys—towards unobstructed access to relevant chemistry information. In The Future of the History of Chemical Information (pp. 127-148).

[19] Graham, S. J.; Hancock, G.; Marco, A. C.; Myers, A. F. 2013. The USPTO trademark case files dataset: Descriptions, lessons, and insights. Journal of Economics & Management Strategy, 22(4), 669-705.

[20] Kattenborn, T.; Leitloff, J.; Schiefer, F.; Hinz, S. 2021. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. ISPRS journal of photogrammetry and remote sensing, 173, 24-49.