# Your Paper Title Here

Wenjie Xu[a,b*], Second Author[a], Third Author[b]

[a]Institutes of Science and Development, CAS, Beijing 100190, China

[b]School of Public Policy and Management, UCAS, Beijing 100049, China

**Abstract** Write your abstract here. The abstract should briefly summarize the main contributions, methodology, and results of your paper. Keep it concise and informative, typically within 150–250 words.

**Keywords:** Keyword 1; Keyword 2; Keyword 3; Keyword 4

## 1 Introduction

Write your introduction here. Provide background context, state the research problem, and outline the paper structure [1].

The main contributions of this paper include:

- Contribution 1
- Contribution 2
- Contribution 3

## 2 Related Work

Review relevant literature and position your work.

## 3 Methodology

### 3.1 Problem Formulation

Describe the problem formulation. Key equations should be numbered:

$$f(x) = \sum_{i=1}^{n} \alpha_i \cdot g_i(x) \tag{1}$$

**Definition 3.1** (Term Name)**.** Provide the definition here.

**Theorem 3.2** (Theorem Name)**.** *State the theorem here.*

**Lemma 3.3** (Lemma Name)**.** *State the lemma here.*

### 3.2 Algorithm

Algorithm 1 shows the main algorithm workflow.

*Corresponding author: xuwenjie@example.edu.cn

**Algorithm 1: Algorithm Name**

**Input:** Input parameters
**Output:** Output result
Step 1 description
**for** *each iteration* **do**
    Update step
**return** *Final result*

表 1: **Dataset Statistics**

| Metric | Dataset A | Dataset B |
|---|---|---|
| Samples | XXX | XXX |
| Features | XXX | XXX |
| Classes | XXX | XXX |

表 2: **Performance Comparison**

| Method | Metric 1 | Metric 2 | Metric 3 |
|---|---|---|---|
| Baseline 1 | XX.X | XX.X | XX.X |
| Baseline 2 | XX.X | XX.X | XX.X |
| **Ours** | **XX.X** | **XX.X** | **XX.X** |

# 4 Experiments

## 4.1 Experimental Setup

Describe datasets, baseline methods, and evaluation metrics.

## 4.2 Results

Table 2 shows the comparison results. Our method outperforms all baselines.

# 5 Conclusion

Summarize the main findings and discuss future work directions.

# Appendix A Supplementary Proofs

This section provides detailed proofs for the theorems presented in the main text.

## A.1 Detailed Proof of Theorem 1

**Theorem A.1** (Convergence Theorem)**.** *Assume the objective function $f(x)$ is L-smooth, and the learning rate $\eta \leq \frac{1}{L}$. Then Algorithm 1 guarantees convergence after $T$ iterations:*

$$\min_{t\in[T]} \|\nabla f(x_t)\|^2 \leq \frac{2(f(x_0) - f^*)}{\eta T} \tag{2}$$

证明. By the definition of $L$-smoothness, for any $x, y$ we have:

$$f(y) \leq f(x) + \nabla f(x)^\top (y - x) + \frac{L}{2}\|y - x\|^2 \tag{3}$$

Let $y = x_{t+1} = x_t - \eta \nabla f(x_t)$, substituting into the above inequality:

$$f(x_{t+1}) \leq f(x_t) - \eta\|\nabla f(x_t)\|^2 + \frac{L\eta^2}{2}\|\nabla f(x_t)\|^2 \tag{4}$$

$$= f(x_t) - \eta\left(1 - \frac{L\eta}{2}\right)\|\nabla f(x_t)\|^2 \tag{5}$$

When $\eta \leq \frac{1}{L}$, we have $1 - \frac{L\eta}{2} \geq \frac{1}{2}$, therefore:

$$f(x_{t+1}) \leq f(x_t) - \frac{\eta}{2}\|\nabla f(x_t)\|^2 \tag{6}$$

Summing over $t = 0, 1, \ldots, T-1$:

$$f(x_T) \leq f(x_0) - \frac{\eta}{2}\sum_{t=0}^{T-1}\|\nabla f(x_t)\|^2 \tag{7}$$

Rearranging and noting that $f(x_T) \geq f^*$ completes the proof. □

# Appendix B Additional Experiments

## B.1 Performance Comparison on Different Datasets

Table 3 presents detailed performance comparisons of our method across multiple public datasets.

**表 3: Performance comparison across datasets (Accuracy %)**

| Method | CIFAR-10 | CIFAR-100 | ImageNet | Average |
|---|---|---|---|---|
| ResNet-50 | 94.2 | 74.5 | 76.1 | 81.6 |
| ResNet-101 | 95.1 | 77.3 | 78.5 | 83.6 |
| ViT-B/16 | 96.2 | 81.4 | 80.2 | 85.9 |
| **Ours** | **97.1** | **83.2** | **82.8** | **87.7** |

## B.2 Hyperparameter Sensitivity Analysis

We conducted sensitivity analysis for key hyperparameters. **??** shows the impact of different learning rates and batch sizes on model performance.

The experimental results demonstrate that the model performance remains relatively stable when the learning rate is in the range $[10^{-4}, 10^{-3}]$ and the batch size is in $[32, 128]$.

# References

[1] First Author and Second Author. A sample paper title. *Journal Name*, 1(1):1–10, 2023.