ISSN: 3104-5235

INNO PRESS

September 2025

Journal of Emerging Applied Artificial Intelligence

Volume 1 / Issue 6

Issue 6 – Foundations of Emerging Applied Artificial Intelligence

The Journal of Emerging Applied AI (JEAAI) is pleased to present its inaugural issue, establishing a dedicated forum for high-quality, peer-reviewed scholarship at the intersection of artificial intelligence theory and real-world application. This first issue reflects the journal's foundational mission: to advance and disseminate research that demonstrates the transformative potential of AI technologies across sectors and disciplines.

This opening volume features contributions that exemplify the journal's emphasis on rigorously developed, practically deployed AI systems. The selected articles cover a spectrum of domains—including healthcare, robotics, transportation, education, and sustainability—demonstrating the breadth of AI's impact when translated from conceptual innovation to applied implementation.

With a commitment to methodological soundness, interdisciplinary relevance, and societal benefit, JEAAI aims to become a leading platform for scholars, practitioners, and innovators who are engaged in solving real-world problems through intelligent systems. The journal's scope encompasses original research, technical reports, case studies, and critical perspectives, all grounded in applicability and reproducibility.

We invite the academic and professional community to engage with JEAAI as contributors, reviewers, and readers, and to join us in shaping a future where applied artificial intelligence drives meaningful and responsible progress.

Editor-in-Chief

Section Editors

Dr. Meng Liu
Research Scientist, NVIDIA

Dr. Liu's research interests include graph neural networks, clustering, and multimodal learning. He has published over 20 papers in leading venues such as *Advanced Science*, *IEEE TPAMI*, *IEEE TKDE*, *CVPR*, *ICML*, and *ICLR*. His work includes an ESI Hot Paper and a Highly Cited Paper, with over 1,000 citations. He has received several awards, including Best Paper at the 2024 China Computational Power Conference and a DAAD AInet Fellowship.

Dr. Zhongbin Luo
Professor-level Senior Engineer, China Merchants Chongqing Communications Research & Design Institute. Master's Supervisor, Chongqing Jiaotong University & Shijiazhuang Tiedao University

Dr. Luo's research focuses on intelligent transportation, traffic safety, and vehicle–road collaboration. He has led over ten national and provincial research projects, holds 11 invention patents, and serves as an expert reviewer for journals such as *IEEE Access* and *PLOS ONE*.

Dr. Ruichen Xu
Postdoctoral Fellow, Department of Civil & Environmental Engineering,University of Missouri, Columbia, USA

Dr. Xu's research interests include hydrological ecology, AI-based flood forecasting, and sediment–pollutant dynamics. He has led or contributed to more than ten projects in China and the U.S. and has published over 20 peer-reviewed papers. He holds patents in environmental monitoring and serves as a reviewer for journals like *Journal of Hydrology* and *Ecological Indicators*.

A/Prof. Jinghao Yang

Assistant Professor, Electrical and Computer Engineering, The University of Texas Rio Grande Valley, USA

Dr. Yang has taught in the U.S. and specializes in applying machine learning to intelligent manufacturing systems. His research bridges intelligent sensing, control, and adaptive design with industrial applications, contributing to smart production technologies and data-driven innovation.

Luxin Zhang
PhD Candidate, School of Engineering, Computer and Mathematical Sciences, Auckland University of Technology, New Zealand

Luxin Zhang is currently pursuing her PhD in Artificial Intelligence. Her research focuses on machine learning algorithms and their applications in intelligent systems. As Managing Editor, she is responsible for manuscript assignment, editorial coordination, and issue scheduling. Based in New Zealand, she serves as a central figure in the journal's daily operations.

Yihan Zhao
PhD Candidate, University of Auckland, New Zealand

Yihan Zhao holds a Master's degree from Peking University and is currently a PhD candidate at the University of Auckland. Her research explores the intersection of communication, culture, and technology, with a focus on how algorithms reshape cultural expression and the subjectivity of marginalized communities. She previously served as an Assistant Research Fellow at the Development Research Centre of the State Council in China, contributing to national research projects. She has curated and coordinated panels for the China Development Forum, facilitating high-level dialogue on AI, sustainability, and governance.

Shen (Jason) Zhan
Graduate Researcher, University of Melbourne, Australia

Jason Zhan holds an Honours degree in Civil and Environmental Engineering from the University of Auckland and is currently a PhD researcher in the Teaching & Learning Lab at the University of Melbourne. He combines industry and academic experience, with a background in structural engineering and teaching. His research focuses on employability assessment and curriculum design in engineering education, with growing interest in the role of AI in authentic assessment and personalized learning.

Contents

A Quantitative Comparison of Large Language Models and Commercial Services for the Translation of Chinese Legal Texts

Fei Qu

School of Foreign Languages, Southwest University of Political Science and Law, Chongqing, China

Abstract—The proliferation of Large Language Models (LLMs) presents transformative potential for professional domains, yet their application in the high-stakes field of legal translation requires rigorous empirical validation. This study conducts a quantitative comparison of the translation quality between two leading LLMs (Gemini 2.5 Pro, ChatGPT 4o) and two reputable commercial translation (CT) services (PKU Law, Wolters Kluwer). The evaluation uses the English translations of the General Provisions of the Criminal Law of the People's Republic of China, with quality assessed through the automated metrics of Bilingual Evaluation Understudy (BLEU) and Translation Edit Rate (TER). Statistical analysis of the four individual sources revealed significant performance differences, with Gemini demonstrating a superior output compared to ChatGPT and, on some measures, PKU Law. However, a subsequent comparison between the aggregated LLM and CT groups found no statistically significant difference in translation quality for either BLEU or TER scores. This study posits that this apparent parity is a methodological illusion that stems from the profound limitations of lexical-based metrics. These metrics reward the superficial fluency of LLMs but are incapable of assessing functional equivalence, thereby failing to penalize critical semantic and legal errors. The findings conclude that despite the impressive coherence of LLM outputs, the nuanced, jurisdiction-specific expertise of human professionals remains the indispensable arbiter of quality and validity in legal translation.

Index Terms—Large Language Models, ChatGPT, BLEU, TER, legal translation

I. INTRODUCTION

The contemporary technological landscape is defined by the rapid development and pervasive integration of Large Language Models (LLMs) across a multitude of professional sectors [1]. These models, representing a significant evolution from earlier paradigms such as Neural Machine Translation (NMT), possess sophisticated capabilities for processing, generating, and interpreting human language. Their advanced architectures enable them to tackle complex informational tasks, driving innovation and transforming workflows in fields as diverse as medicine, finance, and, increasingly, law. The potential of LLMs to automate document drafting, assist in legal research, and provide translation services has generated considerable interest within the legal community.

Within the broader field of language services, legal translation constitutes a uniquely demanding and high-stakes domain. Unlike general-purpose translation, legal translation requires not only linguistic fluency but also profound domain-specific knowledge. The fidelity of a legal translation is paramount, as it must maintain absolute terminological precision, correctly interpret concepts specific to the source and target legal jurisdictions, and preserve the precise legal intent and nuances of the original text. The consequences of error are severe; a single mistranslated term or misinterpreted clause can lead to contractual disputes, regulatory non-compliance, and the invalidation of legal documents in court. Research has documented systematic errors that Al systems make in legal contexts, which underscore these risks. For instance, machine translation models have been observed to mistranslate the legal term "warrant" as the more generic "court order, " a substitution that significantly downplays the legal severity of the document. Similarly, contextual misunderstandings can lead to absurd yet dangerous outputs, such as translating "charged with a battery" as "loaded with a case of batteries". Such errors highlight a critical gap between the general linguistic competence of Al and the specialized precision required for legal practice.

While the fluency and general capabilities of modern LLMs are impressive, their efficacy and reliability for specialized legal translation remain insufficiently quantified. There is a pressing need for empirical evidence that compares the output of these new generative models against the established, human-curated translations provided by professional services, which have long been the standard in the legal industry. This study addresses this gap by conducting a rigorous quantitative analysis of translation quality. The primary objective of this paper is to compare the translation quality of two state-of-the-art LLMs (Gemini 2.5 Pro and ChatGPT 4o) with two reputable commercial legal translation databases (PKU Law and Wolters Kluwer). The evaluation is performed on their respective English translations of the General Provisions of the Criminal Law of the People's Republic of China, using the widely accepted automated metrics BLEU and TER. This objective is operationalized through the following research questions (RQs):

RQ1: Are there statistically significant differences in the translation quality, as measured by BLEU and TER scores, among the four translation sources (Gemini, ChatGPT, PKU Law, Wolters Kluwer) when translating the General Provisions of the PRC Criminal Law?

RQ2: Is there a statistically significant difference in the aggregate translation quality, as measured by BLEU and TER scores, between the Large Language Model (LLM) group and the Commercial Translation (CT) group?

II. Literature Review

A. The Evolution of Automated Translation: From NMT to Large Language Models

The field of automated translation has undergone a profound transformation over the past decade, moving from the paradigm of Neural Machine Translation (NMT) to the current era dominated by Large Language Models (LLMs). NMT, which utilizes deep neural networks to process entire sentences, marked a significant advance over previous statistical methods, substantially improving the fluency and accuracy of machine-generated text. These systems, however, were fundamentally designed as specialized engines optimized for the singular task of translation, trained on curated parallel corpora of source and target sentence pairs [1].

The emergence of LLMs represents not an incremental improvement but a fundamental paradigm shift in artificial intelligence and its application to language [2]. Unlike NMT systems, LLMs are general-purpose models trained on exceptionally vast and diverse text corpora, which endows them with a more comprehensive, context-aware understanding of language and a much broader range of capabilities. This generalist training allows LLMs to adapt to new tasks with minimal or no task-specific data, a technique known as zero-shot or few-shot learning. This versatility has enabled their rapid application across numerous specialized domains, including healthcare and, increasingly, law [3]. In the context of translation, this generalized capability manifests as superior fluency and an enhanced ability to handle context across longer documents, with some models designed to mimic complex human-like processes such as analyzing a source text before rendering a translation.

This technological evolution has introduced a critical tradeoff between fluency and accuracy. The training objective of an LLM, which involves predicting the next word in a sequence based on massive general-domain data, inherently optimizes for linguistic coherence and naturalness. This results in outputs that are exceptionally fluent and conversational. However, this same process makes them susceptible to generating plausible but factually incorrect information, a phenomenon known as hallucination, because their primary goal is linguistic plausibility rather than strict fidelity to a source text. Conversely, NMT systems are trained specifically on parallel texts to optimize for accurate source-to-target mapping, which often makes them more reliable for direct translation accuracy within their trained domains, though their output may be more rigid and less natural-sounding. This inherent tension between the generalist fluency of LLMs and the specialist accuracy of NMT creates a central challenge for evaluation, particularly in high-stakes fields where precision is paramount.

B. The Exigencies of Legal Translation and the Principle of Functional Equivalence

Legal translation constitutes a uniquely demanding and high-stakes domain that magnifies the aforementioned challenges. The task requires not only bilingual fluency but also profound domain-specific knowledge, including an understanding of disparate legal systems and their distinct terminologies [4]. The consequences of error are severe; a mistranslated term or a misinterpreted clause can invalidate contracts, create regulatory liabilities, and subvert judicial outcomes. Therefore, the core of legal translation is the pursuit of absolute precision and consistency in the use of specialized terminology [5,6].

Given these exigencies, the theoretical benchmark for quality in legal translation has long since moved beyond simplistic notions of literalism. Functional equivalence posits that the primary objective of a translator is not to achieve formal, word-for-word correspondence but to produce a target text that has an equivalent effect on its audience. In the legal sphere, this translates to producing a text that has an equivalent legal effect within the target jurisdiction. Achieving this often requires the translator to employ sophisticated strategies such as adaptation, explanation, and the substitution of functionally analogous terms, particularly when a direct conceptual counterpart is absent in the target legal system. This process is an act of cross-jurisdictional communication that demands a deep understanding of the cultural and systemic underpinnings of law in both the source and target contexts. This principle is therefore fundamentally at odds with any evaluation methodology that relies on simple lexical or structural similarity [7,5,6].

C. Evaluating Machine Translation: A Critical Review of Automated Metrics

The practical need for rapid, scalable, and inexpensive evaluation methods has led to the widespread adoption of automated metrics in machine translation research. Human evaluation, while considered the gold standard, is a slow, costly, and subjective process, making it unsuitable for the iterative development cycle of modern MT systems. Metrics such as the Bilingual Evaluation Understudy (BLEU) and Translation Edit Rate (TER) were developed to serve as automated proxies for human judgment [8,9,10].

The BLEU metric is founded on the principle that "the closer a machine translation is to a professional human translation, the better it is". It operates by calculating the precision of matching n-grams (contiguous word sequences) between a candidate translation and one or more human-authored reference translations, applying a brevity penalty to penalize outputs that are too short [9]. TER, conversely, is designed to approximate the post-editing effort required by a human. It calculates the

minimum number of edits, insertions, deletions, substitutions, and shifts, needed to transform a candidate translation to match a reference translation exactly [11].

Despite their widespread use, these lexical-based metrics have been the subject of extensive and long-standing scholarly criticism. The most fundamental critique is that they do not measure translation quality but rather surface-level string similarity to a given reference, a crucial distinction that is often overlooked. Their primary flaw is an inability to account for lexical variation; they cannot recognize synonyms or valid paraphrases, meaning a perfectly accurate translation that uses different wording from the reference is unfairly penalized. This can lead to paradoxical results where nonsensical sentences that happen to contain correct n-grams receive high scores. Consequently, numerous studies have demonstrated a poor correlation between these metrics and human judgments of quality, leading expert bodies to describe them as "artificial and irrelevant for production environments" for over a decade.

This critique becomes particularly salient when applying these metrics to legal translation, creating a profound mismatch between the theoretical requirements of the task and the operational mechanism of the evaluation tool. The theoretical goal of legal translation is to preserve the legal function of a text, which may necessitate altering its linguistic form. The methodological tools, BLEU and TER, operate by rewarding the preservation of linguistic form (lexical and structural matching) and are incapable of assessing legal function. This creates a situation where the more skillfully a system achieves functional equivalence through non-literal but legally correct phrasing, the more likely it is to be penalized. This fundamental conflict highlights why these metrics are theoretically inappropriate for this specific task. Recognizing these deficiencies, the research community has actively developed and transitioned toward semantically-aware metrics based on contextual embeddings, such as BERTScore and COMET, which demonstrate a significantly higher correlation with human judgments by capturing semantic similarity rather than mere lexical overlap [12,8].

D. Identifying the Research Gap: Quantifying LLM Performance in Legal Translation

The existing body of literature reveals a clear and pressing research gap. While LLMs exhibit transformative potential for language tasks, their application to the high-stakes, nuanced domain of legal translation remains a nascent area of inquiry, fraught with documented challenges related to terminological consistency and accuracy. Scholarly consensus indicates a scarcity of empirical studies that rigorously evaluate the performance of state-of-the-art LLMs in this domain, particularly in direct comparison to the established commercial translation services that are the de facto standard in professional legal practice [2].

This study is positioned as a foundational contribution that directly addresses this gap. It provides a crucial, initial quantitative benchmark comparing the outputs of leading LLMs against reputable commercial services for a significant and complex legal text. The selection of BLEU and TER as evaluation metrics is a deliberate methodological choice. While the profound limitations of these metrics are well-established and acknowledged, their use serves a dual purpose in this context. First, it furnishes a baseline measurement using widely understood, albeit dated, metrics, making the results interpretable within the broader history of MT evaluation. Second, and more importantly, it offers a powerful empirical case study that illustrates the "evaluation crisis" and the methodological-theoretical mismatch inherent in applying lexical metrics to a functionally-driven task like legal translation. By quantifying the performance of these distinct system types, fluency-optimized LLMs and professionally curated commercial translations, this research provides the necessary groundwork to motivate and inform future investigations that must adopt more sophisticated, semantically-aware evaluation paradigms to assess generative models in specialized domains.

III. METHODOLOGY

A. Corpus and Translation Sources

The source text for this empirical investigation consists of the General Provisions (Articles 1 through 101) of the Criminal Law of the People's Republic of China. The General Provisions are highly representative for evaluating legal translation quality. They articulate the foundational doctrines of Chinese criminal law, such as definitions of crime, culpability, punishment, and legal defenses, which apply system-wide. Linguistically, this section exemplifies the core features of PRC legislative style, normative modality, complex parataxis, dense definitions, and exception structures, making it a rigorous testbed for assessing semantic precision and syntactic fidelity. Moreover, it foregrounds key translation challenges such as conceptual calibration, liability structure, and penalty conditions. As a result, it offers both doctrinal depth and linguistic generalizability, serving as a valid and methodologically sound corpus for comparative translation studies. Four distinct sources were used to generate English translations of this corpus. These sources were categorized into two groups for comparative analysis as follows:

Large Language Models (LLMs): This group includes two of the most advanced, Gemini 2.5 Pro developed by Google and ChatGPT 4o developed by OpenAI, publicly available generative AI models at the time of the study.

Commercial Translations (CT): This group includes English translations sourced from two highly reputable commercial databases known for providing professional translations of Chinese legal materials. PKU Law is a leading legal information database in China. Wolters Kluwer is a global

Table 1 Descriptive statistics of four translation sources under BLEU (unit: %)

Translation sources	Min.	25% Percentile	Median	75% Percentile	Max.	Range	Mean	Std. Deviation
Gemini	5.538	24.99	31.21	43.99	75.09	69.55	35.03	14.74
ChatGPT	6.058	18.64	27.62	37.25	60.32	54.26	28.37	13.00
PKU Law	4.309	20.75	27.79	39.43	88.53	84.22	31.07	15.93
Wolters Kluwer	5.581	21.49	30.17	43.66	66.60	61.02	32.24	14.33

Table 2 Normality test under BLEU

Translation sources	Kolmogorov-Smirnov			Shapiro-Wilk		
	Test statistic	Degrees of freedom	P Value	Test statistic	Degrees of freedom	P Value
Gemini	.115	101	.002	.965	101	.010
ChatGPT	.068	101	>.010	.976	101	.063
PKU Law	.100	101	.015	.958	101	.003
Wolters Kluwer	.071	101	>.010	.980	101	.120

Table 4 Multiple comparisons of Translation sources under BLEU

Dunn's multiple comparisons test	Mean rank diff.	Significant?	Summary	Adjusted P Value
Gemini vs. ChatGPT	50.47	Yes	*	.013
Gemini vs. PKU Law	34.21	No	ns	.224
Gemini vs. Wolters Kluwer	20.27	No	ns	>.999
ChatGPT vs. PKU Law	-16.26	No	ns	>.999
ChatGPT vs. Wolters Kluwer	-30.20	No	ns	.396
PKU Law vs. Wolters Kluwer	-13.95	No	ns	>.999

provider of professional information, software solutions, and services for the legal and regulatory sectors.

B. Evaluation protocol

The evaluation protocol involved generating or collecting the English translations of all 101 articles from each of the four sources. These 404 translations (101 articles × 4 sources) were then systematically evaluated against the reference translation. Automated scripts were used to compute sentence-level BLEU and TER scores for each article from each source, resulting in a dataset of 101 data points per metric for each of the four translation systems.

C. Procedure

A two-stage statistical analysis was conducted using a significance level (alpha) of $p<.05$ for all inferential tests. All statistical analyses were performed on the provided dataset.

The first stage aimed to answer RQ1 by comparing the performance of the four individual translation sources. The assumption of normality for the score distributions of each source was assessed using the Shapiro-Wilk test, as it is generally more powerful for smaller sample sizes. The Shapiro-Wilk test indicated that the BLEU score data for the Gemini and PKU Law groups significantly deviated from a normal distribution. Consequently, the non-parametric Kruskal-Wallis H test was selected as the appropriate method to compare the median BLEU scores across the four groups. Dunn's test with Bonferroni correction for multiple comparisons was chosen as the post-hoc test to identify which specific pairs of groups differed significantly. The Shapiro-Wilk test indicated that the TER score data for all four groups did not significantly violate the assumption of normality ($p>.05$ for all). Therefore, a one-

way Analysis of Variance (ANOVA) was employed to compare the mean TER scores. Tukey's Honestly Significant Difference (HSD) test was selected for post-hoc pairwise comparisons to determine where the significant differences lay.

The second stage aimed to answer RQ2 by comparing the performance of the LLM group against the CT group. The data from the individual sources were aggregated into the

Table 5 Descriptive statistics of four translation sources under TER (unit: %)

Translation sources	Min.	25% Percentile	Median	75% Percentile	Max.	Range	Mean	Std. Deviation
Gemini	11.11	37.25	49.21	55.56	81.25	70.14	46.55	14.49
ChatGPT	14.29	47.50	55.81	62.79	85.00	70.71	54.28	13.81
PKU Law	4.309	20.75	27.79	39.43	88.53	84.22	31.07	15.93
Wolters Kluwer	6.667	40.37	52.86	60.19	78.31	67.20	50.81	13.85

Table 6 Normality test under TER

Translation sources	Kolmogorov-Smirnov			Shapiro-Wilk		
	Test statistic	Degrees of freedom	P Value	Test statistic	Degrees of freedom	P Value
Gemini	.096	101	.022	.980	101	.126
ChatGPT	.077	101	>.010	.982	101	.197
PKU Law	.051	101	>.010	.987	101	.431
Wolters Kluwer	.072	101	>.010	.986	101	.362

respective groups (LLM: Gemini + ChatGPT, n = 202; CT: PKU Law + Wolters Kluwer, n = 202). The Shapiro-Wilk test was again used to assess the normality of these aggregated distributions. For the BLEU score comparison, the Shapiro-Wilk test revealed that both the LLM group (p=.001) and the CT group violated the assumption of normality. For the TER score comparison, the LLM group also violated normality. Due to these violations, the non-parametric Mann-Whitney U test was selected as the appropriate method to compare the median scores between the LLM and CT groups for both the BLEU and TER metrics.

IV. RESULTS

We conduct the full suite of statistical tests on the BLEU scores, after which the identical testing protocol is applied to the TER scores. Table 1 shows descriptive statistics of four translation tools under BLEU.r

A non-parametric approach was adopted for the subsequent analysis, as the BLEU score data did not meet the assumption of normality ($p<0.05$), as detailed in Table 2. Consequently, a Kruskal-Wallis H test was performed to evaluate for statistically significant differences in the BLEU scores among the four translation tools. The results, presented in Table 3, indicate a significant difference. The p-value was below the 0.05 significance level, leading to the conclusion that a statistically significant variance exists in the translation quality, as quantified by the BLEU metric, across the four translation sources.

To identify the specific sources of the observed variance, we conducted post-hoc pairwise comparisons. We applied Dunn's correction to the significance values to control for the increased risk of a Type I error from conducting multiple tests. The adjusted results, as shown in Table 4, reveal a statistically significant difference only in the comparison between Gemini and ChatGPT. The remaining pairwise comparisons did not achieve statistical significance.

Table 3 Kruskal-Wallis test

P value	0.017
Exact or approximate P value?	Approximate
P value summary	*
Do the medians vary signif.	Yes
Kruskal-Wallis statistic	10.18

Now that the statistical analysis of the BLEU scores is complete, we will perform the same set of tests on the TER scores. Table 5 provides the descriptive statistics for the four translation sources' TER scores.

We began the analysis of the TER scores with a normality test. The analysis of the TER scores commenced with an evaluation of data normality. As Table 6 shows, the Shapiro-Wilk test confirmed that the TER score data for all four sources adhered to the assumption of normality (p>.050 for all).

Because the data were normally distributed, a one-way Analysis of Variance (ANOVA) was the appropriate parametric test to compare the mean TER scores across the four sources.

Table 7 ANOVA results

ANOVA summary	
F	5.699
P Value	.001
P value summary	***

Significant diff. among means (P < 0.05)?	Yes
R squared	.041

The ANOVA test yielded a statistically significant result (F=5.699, p=.001), which is presented in Table 7. This outcome demonstrates that a significant variance exists among the mean TER scores of the four translation sources. The ANOVA result confirms an overall difference, so a post-hoc test was necessary for the identification of specific pairwise differences between the sources. For this purpose, we utilized Tukey's Honestly Significant Difference (HSD) test.

Table 8 Multiple comparisons of Translation sources under TER

Tukey's multiple comparisons test	Mean Diff.	Below threshold?	Summary	Adjusted P Value
Gemini vs. ChatGPT	-7.731	Yes	***	.001
Gemini vs. PKU Law	-6.725	Yes	**	.006
Gemini vs. Wolters Kluwer	-4.263	No	ns	.158
ChatGPT vs. PKU Law	1.006	No	ns	.961
ChatGPT vs. Wolters Kluwer	3.468	No	ns	.325
PKU Law vs. Wolters Kluwer	2.462	No	ns	.623

The results of the Tukey HSD post-hoc analysis are detailed in Table 8. The test identified statistically significant differences in two specific comparisons. The mean TER score for Gemini was significantly lower than that of ChatGPT

Table 9 Normality test between LLMs and CT under BLEU

Translation sources	Kolmogorov-Smirnov			Shapiro-Wilk		
	Test statistic	Degrees of freedom	P Value	Test statistic	Degrees of freedom	P Value
LLMs	.071	202	.015	.976	202	.001
CT	0.073	202	.010	.976	202	.001

Table 10 Normality test between LLMs and CT under TER

Translation sources	Kolmogorov-Smirnov			Shapiro-Wilk		
	Test statistic	Degrees of freedom	P Value	Test statistic	Degrees of freedom	P Value
LLMs	.076	202	.006	.984	202	.019
CT	0.051	202	>.01	.990	202	.195

Table 11 Mann-Whitney test result between LLMs and CT under BLEU

P value	0.8623
Exact or approximate P value?	Approximate
P value summary	ns
Significantly different (P < 0.05)?	No
One- or two-tailed P value?	Two-tailed
Sum of ranks between LLMs and CT	41109, 40701
Mann-Whitney U	20198

Table 12 Mann-Whitney test result between LLMs and CT under TER

P value	0.3033
Exact or approximate P value?	Approximate
P value summary	ns
Significantly different (P < 0.05)?	No
One- or two-tailed P value?	Two-tailed
Sum of ranks between LLMs and CT	39697,42114
Mann-Whitney U	19194

(p=.001) and also significantly lower than that of PKU Law (p=.006). The remaining pairwise comparisons between Gemini and Wolters Kluwer, ChatGPT and PKU Law, ChatGPT and Wolters Kluwer, and PKU Law and Wolters Kluwer did not produce statistically significant differences.

The analysis subsequently progressed to the second stage. This stage addressed the RQ2 and involved a comparison of the aggregated LLM group and the CT group. The evaluation commenced with the aggregated BLEU scores. A normality test was conducted on these two new groups. As Table 9 shows, the Shapiro-Wilk test indicated that both the LLM group (p=.001) and the CT group (p=.001) significantly deviated from a normal distribution. A parallel analytical procedure was then applied to the TER scores. The Shapiro-Wilk test for normality, presented in Table 10, revealed that the LLM group's data were not normally distributed (p=.019), while the CT group's data did not violate the assumption (p=.195).

Because the data violated the assumption of normality, the non-parametric Mann-Whitney U test was the correct statistical method for the comparison of the two independent groups. The results of this test are located in Table 11. The test yielded a p-value of 0.8623, a value that is substantially greater than the 0.05 significance threshold. Therefore, the analysis concludes that no statistically significant difference exists between the median BLEU scores of the LLM group and the CT group.

A non-parametric test is required when the assumption of normality is not met in at least one of the groups, so the Mann-Whitney U test was again employed. Table 12 displays the outcome of this comparison. The resulting p-value was 0.3033, which indicates the difference between the groups are not statistically significant. Consequently, there is no statistically significant difference between the median TER scores of the LLM and CT groups.

V. DISCUSSION

This section provides a comprehensive interpretation of the statistical findings from the preceding analysis. It situates these results within the broader scholarly discourse on machine translation and legal language, critically appraises the methodological framework of the study, and delineates the implications of the findings for theory, practice, and pedagogy. Finally, it proposes a structured agenda for future research to address the limitations identified and to advance the field..

A. Qualitative Analysis Based on BLEU and TE

A qualitative examination of translations that received low BLEU and high TER scores offers concrete evidence of the failures that lexical metrics can detect, even if they cannot capture the full nuance of legal meaning. An analysis confirms that these low scores are not arbitrary penalties for stylistic variation but are markers of substantive semantic and terminological failures.

ChatGPT produced a translation of Article 94 with a BLEU score of 6.05. The source text defines "司法工作人员"(sīfǎ gōngzuò rényuán), which translates to "judicial staff" or "judicial officers," by listing their specific functions: "侦查、检察、审判、监管" (zhēnchá, jiǎnchá, shěnpàn, jiānguǎn), meaning investigation, prosecution, adjudication, and supervision. ChatGPT's translation rendered this as "State functionary," which is a mistranslation. The term "State functionary" is a much broader category in Chinese law and fails to capture the specific functional roles that define a "judicial officer." Furthermore, the translation entirely omitted the enumerated duties, which are the core legal substance of the article. This omission constitutes a critical failure to achieve functional equivalence, as the definition is rendered legally meaningless without them.

PKU Law's translation of Article 3, which scored 4.3, demonstrates a failure in syntactic and semantic fidelity. The source text establishes the principle of legality (nullum crimen, nulla poena, sine lege) [13], a foundational doctrine in criminal law. The reference translation captures the parallel structure and deontic modality ("shall be subjected to... shall not be subjected to") which correctly conveys the mandatory nature of this legal principle. PKU Law's version, "is to be convicted... is not to be convicted," weakens this legal force by using a less definitive grammatical structure. Moreover, its phrasing is syntactically convoluted and less clear than the reference, which impairs the reader's ability to grasp the precise legal rule being articulated.

Similarly, Wolters Kluwer's translation of the same article, which scored 5.58, also fails to convey the correct legal meaning. The translation begins "For acts that are explicitly defined as criminal acts in law," which is a grammatically awkward and imprecise rendering of the original. More significantly, it introduces the term "offenders," which is not present in the source text and presupposes guilt. The Chinese text speaks of "acts"("行为"- xíngwéi), not the individuals committing them. This subtle shift alters the legal focus from the act itself to the actor, which represents a misinterpretation of the legal principle being established. This choice of terminology results in a translation that is not functionally equivalent to the source.

Gemini's translation of Article 74, which discusses the inapplicability of suspended sentences to specific categories of offenders, received a BLEU score of 5.53. The key legal terms in this article are "累犯" (lěifàn), meaning "recidivists," and "缓刑" (huǎnxíng), which translates to "probation" or "suspended sentence." While Gemini correctly translated "recidivists" and "ringleaders of criminal groups," its choice of "suspension of sentence" over "probation" created a lexical divergence from the reference text. Although "suspension of sentence" is a valid translation of "缓刑," the reference translation preferred "probation." This example highlights how even a seemingly minor lexical choice can contribute to a lower BLEU score, even when the translation may be legally acceptable. However, the accumulation of such minor deviations across a text can lead to a significant penalty under a strict lexical matching system.

A qualitative analysis of translations that received high TER scores provides further insight into the specific deficiencies of each system. The TER metric quantifies the post-editing effort

that a human would require, so a higher score indicates a greater number of necessary edits and, consequently, a lower-quality initial translation. An analysis of high-TER examples reveals that the required edits are not merely stylistic; they consistently involve substantive corrections to legal terminology and sentence structure that are essential for achieving functional equivalence.

ChatGPT's translation of Article 99 received a TER score of 85. The source text clarifies that numerical ranges in the law are inclusive. ChatGPT provided a literal translation of "以上" (yǐ shàng) and "以下" (yǐxià) as "above" and "below." These terms are not the standard phrasings that are used in English-language statutes to define inclusive numerical limits. The reference translation uses the correct legal functional equivalents, which are "not more than" and "not less than." The high TER score accurately reflects the significant post-editing effort that is needed. A human editor must replace the literal but functionally incorrect terms with the appropriate legal terminology to ensure the text has the correct legal effect.

The translation of Article 88 from PKU Law, which scored an exceptionally high 92.77, required extensive revision. The translation used the awkward and non-standard term "criminal element" instead of the more precise term "offender." Its sentence structure was convoluted and employed a weak passive voice, for example, "No limitation... is to be imposed." This phrasing fails to convey the direct and binding nature of the legal rule. The reference translation uses the active and definitive statement "the limitation period is not binding." The sheer number of edits that are necessary to correct the terminology, simplify the syntax, and restore the proper legal force justifies the extremely high TER score.

Wolters Kluwer's version of the same article, with a TER score of 78.31, was more competent but still flawed. It used the word "criminal" where "offender" or "suspect" would be more appropriate before a conviction. It also employed the term "dockets the case" as a translation for "立案" (lì'àn). While "docket" is a plausible choice, the reference translation's phrase "filed for investigation" is also a common and clear equivalent. The sentence structure, particularly in the second half of the article, was complex and less direct than the reference. The high TER score reflects the need for these terminological and syntactic adjustments so that the translation aligns with standard legal phraseology.

Gemini's translation of Article 74 received a TER score of 81.25, and this case highlights a specific characteristic of the metric. The translation used "a suspension of sentence" for the Chinese term "缓刑" (huǎnxíng). The reference translation selected "probation." Both terms are conceptually related and can be considered valid translations in different contexts. However, they are not perfect synonyms. The high TER score resulted almost entirely from this single major terminological substitution, along with minor grammatical shifts. This example demonstrates that TER heavily penalizes a translation when it deviates from the specific lexical choices of the single reference text, even if the chosen alternative is functionally similar.

B. Interpretation of Principal Findings: LLM Fluency and the Illusion of Statistical Parity

The statistical analysis yielded a bifurcated set of results that demand careful interpretation. The RQ1 investigated the performance of the four individual translation sources. The findings revealed statistically significant differences among them, with Gemini demonstrating a notable performance advantage over ChatGPT, reflected in both a higher median BLEU score (p=.013) and a significantly lower mean TER score (p=.001). Furthermore, Gemini's output required significantly fewer edits than that of PKU Law, as indicated by its lower mean TER score (p=.006).This suggests that on a structural level, Gemini's translations were closer to the reference text than those from a reputable commercial database, a testament to the rapid, iterative advancements in the architecture of leading LLMs.

However, the analysis for RQ2, which compared the aggregated LLM group against the CT group, produced a starkly different outcome. The results of the Mann-Whitney U tests indicated no statistically significant difference between the two groups for either the BLEU metric (p=.8623) or the TER metric (p=.3033). A superficial reading of this finding might suggest that the translation quality of state-of-the-art LLMs has achieved parity with established, professional legal translation services. This discussion posits that such a conclusion is not only premature but is likely an artifact of the evaluation methodology itself. The observed statistical parity is, in effect, a methodological illusion that reveals more about the inherent limitations of the chosen metrics than it does about the true comparative quality of the translations.

This illusion of parity can be deconstructed by examining the interplay between the known strengths and weaknesses of both the LLMs and the evaluation metrics. LLMs are engineered to generate text that is exceptionally fluent and grammatically coherent, a capability derived from their training on vast text corpora. This inherent fluency naturally results in a high degree of n-gram overlap with any well-formed reference text, thereby inflating BLEU scores. Conversely, these same models are known to struggle with domain-specific knowledge and contextual nuance, leading to systematic errors in specialized fields such as law. These errors, which can be semantically and legally catastrophic, are often not adequately penalized by lexical metrics. For instance, an LLM might produce a highly fluent sentence that contains a critical terminological error, while a professional translation might use different phrasing, functionally correct but lexically divergent from the single reference, thereby receiving a comparatively lower BLEU score. Across a large corpus, the LLMs' high scores for fluency can effectively mask their low scores for accuracy, while the professional translations' perfect accuracy may be penalized for lexical divergence. The net effect is the potential cancellation of these differences, leading to the non-significant result observed in the group comparison. Therefore, the finding of no difference does not signify equivalent quality; rather, it highlights a fundamental misalignment between the evaluation tools and the complex nature of legal translation.

C. The Inadequacy of Lexical Metrics for Assessing Functional Equivalence in Legal Translation

The central challenge in evaluating legal translation quality lies in defining an appropriate theoretical standard. The field has long moved past simplistic notions of literalism, converging instead on the principle of "functional equivalence" as the benchmark for high-quality translation. Advanced by theorists, this principle holds that the primary objective of a legal translator is not to achieve formal, word-for-word correspondence, but to produce a target text that has an equivalent legal effect within the target jurisdiction. This is a process of cross-jurisdictional communication that involves translating legal concepts and their intended consequences, an act that often requires adaptation, explanation, and the use of functionally analogous terms rather than literal equivalents.

This study's findings bring the profound conflict between this theoretical standard and the chosen evaluation metrics into sharp relief. The BLEU metric, by its very design, measures n-gram precision and is thus a proxy for formal, not functional, equivalence. It is fundamentally incapable of recognizing legitimate synonyms or valid paraphrasing, which are indispensable strategies for a translator attempting to achieve functional equivalence when a direct terminological counterpart is absent in the target legal system. Similarly, the TER metric measures the post-editing effort required to make a translation match a reference. However, it does so by counting edits without differentiating their severity; a single-word substitution that corrects a fundamental legal misinterpretation is weighted no more heavily than a trivial edit of punctuation or style. Consequently, TER quantifies the effort of correction but fails to capture the significance of the errors being corrected.

This misalignment creates what can be termed a fluency trap. Because modern LLMs excel at producing fluent text, they generate outputs that appear to be of high quality when assessed by metrics that prioritize fluency and lexical similarity. This masks deep-seated failures in achieving the functional equivalence that is the cornerstone of legal validity. This phenomenon poses a significant risk, as non-expert users or automated evaluation pipelines could erroneously conclude that raw LLM output is a viable substitute for professional legal translation, a conclusion that this study's results, when properly contextualized, strongly refute.

D. Implications for Legal Practice, AI Development, and Translation Pedagogy

This study carries substantial implications for key stakeholder groups. The nuanced interpretation of statistical parity between LLMs and commercial services offers valuable insights for legal professionals, AI developers, and translation educators alike.

For legal professionals, the findings offer a critical caution: the appearance of parity in automated metrics can be misleading. LLMs such as Gemini, while fluent, remain prone to domain-specific semantic errors that pose risks in high-stakes contexts. This underscores the indispensability of human oversight. The most responsible application of LLMs in legal translation lies within a Machine Translation Post-Editing (MTPE) workflow, where AI-generated drafts are subject to expert review. Effective MTPE requires not only legal and linguistic competence, but also high-quality source texts, clear editorial standards, and well-maintained legal termbases to guide the post-editing process.

For AI developers and the MT research community, the study exposes the limitations of relying on traditional metrics like BLEU and TER, which privilege surface fluency over legal adequacy. Continued reliance on such metrics risks incentivizing models that perform well numerically but fail semantically. Progress in this field demands the development of domain-specific evaluation protocols, including high-quality multi-reference test sets for legal texts and the adoption of metrics better aligned with human judgments of legal accuracy. Moreover, the results suggest that substantial improvements in legal translation quality will likely come through domain-adaptive fine-tuning on curated legal corpora that equip models with the specialized knowledge they currently lack.

For translation pedagogy, the study signals a pressing need to update training programs. Legal translators of the future must possess critical AI literacy—understanding both the capabilities and the systemic limitations of LLMs. MTPE training should be a core component of the curriculum, emphasizing not only grammatical correction but the ability to detect and resolve subtle, high-stakes legal errors. Additionally, students must be equipped to critically assess automated evaluation metrics, recognize their limitations, and make informed decisions about tool adoption in professional contexts.

E. Methodological Limitation

Beyond the theoretical shortcomings of the chosen metrics, this study faces several methodological limitations that constrain the interpretation and generalizability of its findings.

The most critical limitation lies in the exclusive reliance on automated evaluation metrics, particularly BLEU and TER, which have been widely criticized in recent scholarship. Studies have consistently shown that these metrics correlate poorly with human judgments, especially when comparing high-performing systems or evaluating legal translations at the sentence level. They fail to capture semantic adequacy, overemphasize surface-level lexical overlap, and rely heavily on a single reference translation. As a result, LLMs, which often produce lexically diverse yet semantically accurate output, may be systematically underrated. Furthermore, these metrics lack diagnostic power; they provide no insight into the nature of errors, whether lexical, syntactic, or terminological.

Another key constraint is the single-reference bottleneck. Evaluating translations against only one reference text penalizes legitimate variation, particularly in legal translation where multiple phrasings may convey equivalent legal meaning. This rigid comparison can obscure the true strengths of systems capable of producing functionally accurate but differently worded outputs.

The scope of the study also limits generalizability. The analysis was restricted to one document and to a single

language pair: Chinese-to-English. While this section offers important doctrinal and linguistic challenges, its terminological focus may not represent other legal domains such as contract, civil, or administrative law, which follow different conventions. Additionally, LLM performance varies across language pairs, particularly for low-resource languages, further limiting extrapolation.

Acknowledging these limitations is essential for interpreting the findings in context and for informing future research on evaluation strategies that better reflect the semantic and functional demands of legal translation.

F. Directions for Future Research

The limitations identified in this study give rise to a clear and structured agenda for future research. To build upon these preliminary findings and to arrive at more robust and valid conclusions, subsequent investigations should proceed along four primary avenues.

First, the most critical next step is to conduct a parallel study that incorporates human-centric evaluation methodologies to triangulate the current findings. This would provide the "gold standard" assessment that is currently missing. Such a study could employ a multi-faceted approach, including Direct Assessment (DA), where bilingual legal experts rate the translations from all four sources on a continuous scale for adequacy and fluency. This should be complemented by a detailed error typology analysis, in which human annotators manually identify and categorize the errors made by each system (e.g., terminological, syntactic, semantic, omission). It is hypothesized that such an analysis would reveal significant differences in the types and severity of errors between the LLM and CT groups, thereby dismantling the illusion of parity produced by the lexical metrics.

Second, the existing corpus of translations should be re-evaluated using modern, semantically-aware automated metrics. Research has shown that metrics such as COMET and BERTScore, which are based on contextual embeddings from pretrained language models, correlate much more highly with human judgments of quality than BLEU or TER. A re-analysis using these advanced metrics would directly test the central hypothesis of this discussion: that superior evaluation tools will detect a statistically significant quality difference between the LLM and CT groups that the legacy metrics failed to capture. Such a finding would provide strong empirical support for the machine translation community's shift away from BLEU.

Third, future research should investigate the impact of domain-specific adaptation on LLM performance. The LLMs used in this study were general-purpose models. A powerful comparative study would involve fine-tuning a model like Gemini or ChatGPT on a large, high-quality corpus of parallel Chinese-English legal texts. A subsequent evaluation comparing the output of this fine-tuned model against the general-purpose model and the commercial services could quantify the performance gains attributable to domain specialization and determine whether such adaptation can genuinely close the quality gap with professional, human-led translation workflows.

Finally, the scope of the investigation must be broadened to enhance the generalizability of the findings. Future studies should include a wider variety of legal documents, such as contracts, judicial decisions, and patent applications, which present different linguistic and conceptual challenges. The analysis should also be extended to other language pairs, including low-resource languages, to assess the robustness of LLM performance across different linguistic contexts. An additional and highly practical avenue of research would be to measure the post-editing effort (in terms of time and cost) required to bring the output from each source to a publishable, commercially acceptable standard, which would provide a more pragmatic measure of each system's utility in a real-world professional setting.

To synthesize the core argument of this discussion and to guide future methodological choices, the following table compares the three primary classes of evaluation methodologies.

VI. CONCLUSION

This study's quantitative comparison of LLM and commercial translation services for a specialized legal text yields a paradoxical result. The statistical findings, particularly the lack of a significant difference between the aggregated LLM and CT groups, do not demonstrate that generative AI has achieved parity with professional, human-centric translation. Instead, these results serve as a powerful illustration of the profound inadequacy of purely quantitative, lexical-based evaluation metrics for the high-stakes domain of legal translation. The investigation reveals that the tools commonly used to measure translation quality are fundamentally misaligned with the theoretical and practical requirements of the task, rewarding superficial fluency while remaining blind to critical errors in legal meaning and function.

The rapid advancement in the fluency of models like Gemini is undeniable and represents a significant technological achievement. However, this study concludes that the nuanced, context-aware, and jurisdiction-specific expertise that defines professional legal translation remains an exclusively human capability. For the foreseeable future, the human legal expert is not merely a participant in the quality assurance process but remains the indispensable and definitive arbiter of validity and quality in legal translation.

REFERENCES

[1] S. M. Abdelhalim, A. A. Alsahil, and Z. A. Alsuhaibani, "Artificial intelligence tools and literary translation: A comparative investigation of ChatGPT and Google translate from novice and advanced EFL student translators' perspectives," Cogent Arts and Humanities, vol. 12, no. 1, pp. 1–20, 2025.

[2] M. Bajcic and D. Golenko, "Applying large language models in legal translation: The state-of-the-art," International Journal of LANGUAGE & LAW, pp. 171–196, 2024.

[3] X. Xuan et al., “TransLaw: Benchmarking large language models in multi-agent simulation of the collaborative translation,” arXiv, 2025. [Online]. Available: http://arxiv.org/abs/2507.00875v1.
[4] M. Bajcic and D. Golenko, “Applying large language models in legal translation: The state-of-the-art,” International Journal of LANGUAGE & LAW, pp. 171–196, 2024.
[5] C. Aliyev, “Functional equivalence in the translation of legal terms: Insights from Arabic and Azerbaijani,” vol. 1, no. 4, pp. 162–168, 2025.
[6] J. Huang and G. Wang, “A study on translation strategies of legal English texts from the perspective of functional equivalence theory,” Frontiers in Humanities and Social Sciences, vol. 4, no. 7, pp. 37–43, 2024.
[7] J. Jiang and Y. Zhi, “Translating euphemisms: Analyzing functional equivalence theory in context,” US-China Foreign Language, vol. 22, no. 5, pp. 270–275, 2024.
[8] S. Lee et al., “A survey on evaluation metrics for machine translation,” Mathematics, vol. 11, no. 4, pp. 1–22, 2023.
[9] K. Papineni et al., “Bleu: A method for automatic evaluation of machine translation,” in Proc. 40th Annual Meeting on Association for Computational Linguistics, Morristown, NJ, USA, 2001, pp. 1–9.
[10] C. Wu et al., “Statistical machine translation for biomedical text: Are we there yet?,” in AMIA ... Annual Symposium proceedings, 2011, pp. 1290–1299.
[11] M. Snover et al., “A study of translation edit rate with targeted human annotation,” in Proc. 7th Conference of the Association for Machine Translation in the Americas, 2006, pp. 223–231.
[12] T. Zhang et al., “BERTScore: Evaluating text generation with BERT,” in International Conference on Learning Representations, 2020, pp. 1–43. [Online]. Available: http://arxiv.org/abs/1904.09675v3.
[13] B. Xue, The compact English-Chinese dictionary of Anglo-American law. Beijing, China: Peking University Press, 2013.

From Detection to Prediction: A Multimodal Deep Learning Framework for Proactive Fall Risk Monitoring in Smart Aging

Haoze Ni[1], Xinyue Huang [2], Wuyang Zhang[3]

[1]College of Communication , Emerging Media Studies(EMS), Boston University, Boston, United States

[2]Independent researcher, New York, United States

[3]Department of Electrical and Computer Engineering, University of Massachuetts, Amherst , United States

Abstract

artificial intelligence has seen widespread adoption across diverse domains, and its potential in smart aging warrants further exploration[11, 16, 9, 7]. Falls are a leading cause of morbidity and mortality among oli2024deeplder adults, with substantial social and economic impact. Existing fall detection systems primarily operate in a reactive manner, recognizing incidents only after they occur. While useful, such approaches do not prevent injury and often suffer from low adherence or high false alarm rates. In this work, we propose a predictive framework for *in-home fall risk assessment* that shifts the focus from post-event detection to pre-event forecasting. Our system integrates multimodal sensing—including wearable inertial measurement units (IMUs), millimeter-wave radar, and pressure sensors—with a temporal deep learning architecture trained via self-supervised pretraining and personalized adaptation. By analyzing gait instability, postural transitions, and near-fall events as precursors, the model outputs both a continuous risk score and a probability of falls within multiple future horizons.

Extensive experiments on a naturalistic longitudinal dataset and a public benchmark demonstrate that our approach achieves earlier and more reliable predictions than rule-based, classical machine learning, and purely supervised deep learning baselines. Compared to existing detectors, our system improves AUROC and lead-time while reducing daily false alarms, offering actionable early warnings. Importantly, attention-based interpretability highlights clinically relevant precursors, enhancing trust and adoption in elder care. This work represents a step toward proactive, personalized, and privacy-preserving fall prevention, supporting independent living for the aging population.

Index Terms—Fall prediction, smart aging, multimodal sensing, time-series analysis, self-supervised learning, elder care, deep learning.

1 Introduction

Falls are one of the most severe threats to the health and independence of older adults. According to the World Health Organization, nearly one third of adults aged over 65 experience at least one fall per year, and falls remain the leading cause of injury-related hospitalization and death in this population. Beyond the immediate physical harm, fear of falling contributes to reduced mobility, social isolation, and a diminished quality of life. With the rapid growth of the aging population worldwide, effective solutions for fall prevention and timely intervention are urgently needed.

Artificial intelligence and sensing technologies have increasingly been explored in related domains, ranging from personalized pedometer and gait analysis with wearable IMUs[2], to privacy-preserving healthcare infrastructures for rehabilitation[3], and deep learning approaches for movement and performance analysis in sports[4]. These advances demonstrate the broader potential of AI-driven multimodal analytics for capturing subtle biomechanical patterns, motivating their application in the context of fall risk monitoring and smart aging.

Over the past decade, a large body of research has focused on *fall detection*, leveraging wearable sensors, cameras, or environmental devices to recognize falls after they occur[14, 1, 12]. While detection is valuable for emergency response, it remains inherently reactive: injuries have already taken place by the time an alarm is triggered. Furthermore, practical deployment faces multiple challenges. Wearable solutions require consistent adherence, which is often low among frail individuals. Vision-based methods introduce privacy concerns and are sensitive to occlusion and lighting. Threshold-based or single-sensor systems often suffer from high false alarm rates, undermining user trust and adoption[6, 15].

These limitations motivate a paradigm shift: moving from *detection* to *prediction*[13, 8]. Rather than identifying falls post hoc, the goal is to anticipate elevated risk before an incident occurs. Early warnings would allow caregivers to implement preventive measures such as physical therapy, mobility aids, or environmental adjustments, potentially averting severe outcomes. However, fall prediction is a considerably more challenging task, requiring models that can extract subtle and temporally extended precursors of instability from complex, multimodal data.

In this paper, we introduce a predictive framework for in-home fall risk assessment. Our approach integrates multimodal sensing, including wearable IMUs, radar, and pressure

sensors, with a temporal deep learning model based on Transformer encoders. To overcome the scarcity of fall labels, we incorporate self-supervised pretraining using masked modeling and contrastive objectives. To address inter-individual variability, we employ lightweight personalization strategies that adapt the model to each user's daily patterns. The system outputs both a continuous risk trajectory and binary event probabilities for multiple horizons, providing interpretable and actionable early warnings.

We validate our approach on a longitudinal in-home dataset and a public benchmark. Results demonstrate that our model improves AUROC, AUPRC, and prediction lead-time while significantly reducing false alarms. Contributions of this work are threefold: (i) a novel multimodal predictive framework that shifts from fall detection to fall forecasting; (ii) a self-supervised and personalized architecture for robust learning under scarce and imbalanced labels; and (iii) extensive evaluation showing earlier and more reliable risk prediction with interpretable insights. Together, these advances highlight the potential of predictive fall monitoring as a cornerstone of smart aging technologies.

2 Methods

2.1 Problem Formulation

Let $\mathcal{P}$ denote the set of study participants and $\mathcal{M}$ the set of sensing modalities (e.g., wearable IMU, millimeter-wave radar, pressure arrays, and ambient context sensors). For a given participant $p \in \mathcal{P}$ and modality $m \in \mathcal{M}$, we observe a univariate or multivariate time series $X_p^{(m)}(t) \in \mathbb{R}^{d_m}$ sampled at time t. After synchronization and resampling (Section 2.5), streams are aligned on a common timeline and concatenated to yield a unified multivariate sequence

$$\mathbf{X}_p(t) = \left[X_p^{(1)}(t) \,\|\, X_p^{(2)}(t) \,\|\, \cdots \,\|\, X_p^{(|\mathcal{M}|)}(t)\right] \in \mathbb{R}^D,$$

where $D = \sum_m d_m$. We operate on sliding windows of length W seconds, indexed by their right endpoint t; the model takes as input $\mathbf{X}_{p,t-W:t} \in \mathbb{R}^{L\times D}$, where L is the number of resampled frames in the window. The objective is to predict a continuous, well-calibrated *fall risk* score $r_p(t) \in [0,1]$ reflecting the likelihood of a fall or near-fall in a prospective horizon $[t, t+\Delta]$, together with an event probability $\hat{y}_p(t,\Delta) \in [0,1]$ for the same horizon. We thus learn a mapping

$$(r_p(t), \hat{y}_p(t,\Delta)) \;=\; f_\theta\big(\mathbf{X}_{p,t-W:t}, \mathbf{Z}_{p,t-W:t}\big), \qquad (1)$$

where $\mathbf{Z}_{p,t-W:t}$ denotes auxiliary behavioral descriptors derived from $\mathbf{X}$ (Section **??**). The binary supervision $y_p(t,\Delta) \in \{0,1\}$ indicates whether a clinically-relevant event (fall or near-fall) occurs within $[t, t+\Delta]$. We consider multiple horizons $\Delta \in \{\Delta_1, \Delta_2, \Delta_3\}$ (e.g., 1h, 6h, 24h) to support both imminent-warning and day-ahead risk stratification within a unified formulation.

2.2 Cohort, Setting, and Sensing Modalities

Participants are enrolled under informed consent for longitudinal, in-home monitoring. The sensing configuration balances fidelity, burden, and privacy by combining minimally obtrusive wearables with contactless environmental sensors. A waist- or wrist-worn IMU provides triaxial accelerometry and gyroscopy at native rates of 50–100 Hz, enabling fine-grained gait and posture dynamics. One or more short-range millimeter-wave radars are positioned to cover habitual activity zones (e.g., bedroom, bathroom, corridor); their micro-Doppler returns and velocity fields capture whole-body and limb-specific kinematics without collecting identifiable imagery. Pressure sensors are deployed as floor mats or bed mats to record weight shifts, stance phases, and nocturnal postural changes. Low-duty ambient context sensors, including Bluetooth or UWB beacons and door contacts, provide coarse location and activity cues that enrich temporal context at negligible privacy cost. All streams are buffered and processed on an edge gateway situated in the home; raw audio and video are not persisted or transmitted off-device.

2.3 Event Definitions and Labeling Strategy

The primary endpoint is a medically significant fall as reported by the participant, caregiver, or clinical staff, time-stamped via incident reports or phone follow-ups. Because falls are rare, we expand supervision using *near-fall* proxies drawn from high-jerk perturbations, corrective stepping, abrupt angular velocity peaks, and radar-inferred loss-of-balance patterns that do not culminate in ground impact. Formally, let $\mathcal{E}_p$ be the set of annotated events for participant p, where each event $e \in \mathcal{E}_p$ has an onset time t_e and a type label in $\{\text{fall}, \text{near-fall}\}$. For a prediction window ending at time t, the binary target is defined as

$$y_p(t,\Delta) \;=\; \mathbb{I}(\exists\, e \in \mathcal{E}_p \text{ s.t. } t \le t_e < t+\Delta)\,, \qquad (2)$$

optionally weighting near-falls with a scalar $\alpha \in (0,1]$ in the learning objective to reflect their lower severity while preserving their predictive value. To mitigate annotation jitter, we allow a tolerance ε (e.g., ± 2 min) around t_e when linking events to windows.

2.4 Data Acquisition and Synchronization

All devices publish time-stamped packets to the edge gateway over low-latency links. Clock drift is bounded using network time protocol synchronization at deployment and opportunistic beacon-based re-alignment daily. Streams are resampled to a common frequency f_s (10 Hz unless stated otherwise) via polyphase decimation or band-limited interpolation, depending on the native rate. We resolve inter-stream offsets by maximizing cross-correlation between modality-specific motion energy envelopes over a short calibration procedure at installation and by maintaining a low-order affine correction thereafter. The output is a set of co-registered frames $\{\mathbf{X}_p(t_k)\}_{k=1}^{L}$ per window.

2.5 Preprocessing and Denoising

Preprocessing aims to suppress sensor noise while preserving biomechanically salient dynamics. IMU signals are passed through a zero-phase Butterworth low-pass filter with cutoff f_c adapted to gait cadence (typically $f_c \in [10, 15]$ Hz), after removing gravitational components by quaternion-based orientation estimation or high-pass filtering when attitudes are unavailable. Radar returns are converted into time–frequency representations using a short-time Fourier transform with Hamming windows, and stationary clutter is attenuated through background subtraction and magnitude gating; we further suppress spurious micro-Doppler streaks via median filtering in the time–frequency plane. Pressure arrays are denoised by spatial median filters followed by row/column detrending to account for slow baseline drift. Ambient context signals are encoded as piecewise-constant states and subsequently one-hot or embedded as low-dimensional vectors. All channels are standardized per participant using robust z-scoring (median and interquartile range) to reduce inter-person variability without distorting heavy-tailed motion distributions. Missing samples within a window are imputed using zero-order hold if gaps are shorter than δ seconds, or masked explicitly when longer, enabling the downstream model to remain well-posed under intermittent dropouts.

2.6 Segmentation and Windowing

Continuous streams are partitioned into overlapping windows of length W with stride S (e.g., $W \in [60, 300]$ s and $S \in [10, 30]$ s). This choice captures sufficient context for estimating stability and postural control while enabling near-real-time updates. For multi-horizon prediction, each window is associated with targets $\{y_p(t, \Delta_j)\}_j$ as in (2). To avoid information leakage in temporal evaluation, windows whose horizons overlap an event are assigned exclusively to either training or testing depending on the event time relative to the split boundary.

2.7 Model Architecture

The predictive model f_θ is designed as a multimodal, temporal deep network that jointly exploits raw sensor streams and handcrafted descriptors. Its design reflects three goals: (i) to capture short- and long-range temporal dependencies that underpin gait stability and pre-fall dynamics; (ii) to integrate heterogeneous modalities with complementary signal characteristics; and (iii) to provide both discrete event likelihoods and continuous risk trajectories.

Multimodal Encoders. Each sensing modality is first processed by a dedicated encoder tailored to its signal structure. Wearable IMU sequences are passed through a stack of one-dimensional temporal convolutional layers followed by bidirectional LSTM layers, yielding embeddings that capture both local fluctuations and recurrent dynamics. Radar time–frequency spectrograms are encoded via a lightweight two-dimensional CNN followed by a Transformer encoder, which attends to salient micro-Doppler streaks and velocity bursts. Pressure maps are projected through a spatial CNN with small receptive fields and temporal pooling to summarize stance patterns. Ambient context embeddings are produced by an embedding layer followed by gated recurrent units. The outputs are temporally aligned and projected into a common latent space.

Cross-Modal Fusion. To exploit complementarities while preserving modality-specific nuances, we employ a cross-attention fusion module. Given modality-specific embeddings $\{\mathbf{h}_{1:L}^{(m)}\}_{m=1}^{M}$, the fusion layer computes for each timestep

$$\tilde{\mathbf{h}}_t = \sum_{m=1}^{M} \alpha_t^{(m)} W^{(m)} \mathbf{h}_t^{(m)}, \quad \alpha_t^{(m)} = \mathrm{softmax}_m \left(q_t^\top K^{(m)} \mathbf{h}_t^{(m)} \right),$$

where q_t is a shared query derived from the concatenation of all modalities at t, $K^{(m)}$ is a learned projection, and $W^{(m)}$ is a modality-specific linear transform. This attention-based scheme highlights whichever modality most strongly explains instability at a given moment (e.g., radar for balance loss, IMU for step variability).

Temporal Representation Learning. The fused sequence $\{\tilde{\mathbf{h}}_t\}_{t=1}^{L}$ is passed through a Transformer encoder with multi-head self-attention to capture long-range dependencies and repeated instability motifs. Positional encodings ensure temporal ordering is preserved. To handle windows with partial dropouts, modality-dropout masks are concatenated to the embeddings, allowing the Transformer to learn robustness to missing channels.

Self-Supervised Pretraining. Before supervised training, the encoders are pretrained on unlabeled data using two auxiliary objectives: (i) masked sequence modeling, in which randomly masked segments of $\tilde{\mathbf{h}}_t$ are reconstructed from context, and (ii) temporal contrastive learning, in which neighboring windows from the same subject are treated as positive pairs and distant windows or different subjects as negatives. The resulting loss,

$$\mathcal{L}_{SSL} = \mathcal{L}_{\text{mask}} + \mathcal{L}_{\text{contrast}},$$

encourages invariances that transfer to the rare event prediction task.

Prediction Heads. From the temporal encoder's output, we derive two heads. The classification head applies global average pooling followed by a sigmoid unit to estimate $\hat{y}_p(t, \Delta)$, the probability of a fall/near-fall within the horizon Δ. The regression head uses an attention-weighted pooling followed by a linear layer to estimate the continuous risk score $r_p(t)$. The joint loss is

$$\mathcal{L} = \lambda_1 \,\mathrm{BCE}(y_p, \hat{y}_p) + \lambda_2 \,\mathrm{MSE}(r_p, \hat{r}_p) + \lambda_3 \,\mathcal{L}_{SSL}, \quad (3)$$

with $\lambda_1, \lambda_2, \lambda_3$ balancing classification, regression, and self-supervised objectives.

Personalization and Online Adaptation. Because gait and balance characteristics vary widely across individuals, we implement personalization strategies. During deployment, the model undergoes lightweight subject-specific fine-tuning using a few days of data, updating only normalization parameters and prediction heads. Online calibration is performed with temperature scaling to ensure probability calibration for each subject. In addition, drift detectors such as ADWIN monitor distributional shifts; when triggered, the system either fine-tunes on recent data or reverts to the population model.

2.8 Imbalanced Learning and Weak Labels

Falls are rare events, leading to severe class imbalance. We address this with several strategies. First, near-falls are incorporated as weak positives, weighted by $\alpha < 1$ in the loss to reflect their lower severity yet predictive value. Second, focal loss is applied within the classification head to emphasize hard and minority examples. Third, pseudo-labeling with high-confidence predictions augments the training set in a semi-supervised fashion, gradually refining decision boundaries.

2.9 Training Protocol

We evaluate models under both leave-one-subject-out (LOSO) and temporal hold-out protocols to test cross-person generalization and longitudinal robustness. Windows are sampled with stride S to balance temporal resolution and computational cost. Optimization uses AdamW with a OneCycle learning rate schedule and mixed precision. Data augmentation includes time warping, jittering, Gaussian noise injection, and modality dropout. Multiple prediction horizons ($\Delta = 1\,\text{h}$, $6\,\text{h}$, $24\,\text{h}$) are learned in a multitask manner by adding separate classification heads that share the encoder.

2.10 Inference and Alerting

At runtime, windows are streamed through the model in overlapping fashion, and risk scores are updated every S seconds. To mitigate false alarms, an alert is issued only if $r_p(t) > \tau$ for k consecutive windows. Thresholds are context-aware, with lower τ in high-risk environments such as bathrooms or stairways. Alerts are tiered: green for normal, yellow for moderate risk prompting preventive advice, and red for imminent risk that triggers caregiver notifications. This design integrates predictive analytics into a clinically actionable framework.

2.11 Evaluation Metrics

We assess performance on multiple levels. Event-level metrics include precision, recall, F1, AUROC, AUPRC, and mean prediction lead-time at fixed recall. Calibration quality is quantified with Brier score and Expected Calibration Error. User-level utility is measured as false alarms per day. Ablation studies compare modality subsets, pretrained versus randomly initialized encoders, and personalized versus generic models. Statistical significance is evaluated with bootstrapped confidence intervals and DeLong's test for AUROC differences.

2.12 Explainability and Deployment

To enhance trust and adoption, attention heatmaps identify the temporal regions driving high risk predictions, and SHAP values attribute contributions of handcrafted descriptors. Weekly summaries provide clinicians with interpretable patterns such as rising gait variability or increasing nocturnal instability. For deployment, all processing runs on an edge gateway with federated learning to update shared weights. Differential privacy ensures uploaded gradients are noise-perturbed, and model compression enables execution on embedded NPUs with real-time latency.

3 Experiments and Results

3.1 Datasets

To evaluate the proposed predictive fall risk framework, we conduct experiments on two datasets.

In-house longitudinal dataset. All data were independently and exclusively collected by the authors. We constructed a continuous multimodal dataset. The sensing configuration consisted of a waist-worn IMU (100,Hz), two ceiling-mounted mmWave radars (20,Hz), and pressure mats installed near beds and bathrooms (10,Hz). Ground-truth fall events ($n = 22$) were annotated by caregivers and verified against formal incident reports, while near-falls ($n = 134$) were identified through manual inspection of radar sequences in conjunction with caregiver logs.

3.2 Baselines

We compare our method against representative approaches:

- **Threshold-based IMU detector:** peak acceleration and angular velocity rules with fixed thresholds, as in [**?**].
- **Classical ML:** handcrafted features from IMU signals fed into Random Forests and SVMs.
- **Deep CNN-LSTM:** end-to-end supervised model trained on raw IMU windows without self-supervision or multimodal fusion.
- **Vision-based fall detector:** 3D CNN trained on RGB videos (available only in SisFall).

These baselines represent current paradigms: simple rules, shallow ML, deep supervised learning, and vision-based methods.

3.3 Evaluation Protocol

We employ two complementary validation strategies:

1. **Leave-One-Subject-Out (LOSO):** models trained on all but one subject and evaluated on the held-out subject, rotating across participants. This tests cross-subject generalization.

2. **Temporal hold-out:** for each subject, the first 70% of time windows are used for training and the last 30% for testing. This tests longitudinal robustness.

Prediction horizons $\Delta \in \{1\,\text{h}, 6\,\text{h}, 24\,\text{h}\}$ are evaluated. We report mean performance across subjects.

3.4 Metrics

- **Event-level:** Precision, Recall, F1-score, AUROC, and AUPRC.
- **Lead-time:** average early warning time (minutes) at 80% recall.
- **Calibration:** Brier score and Expected Calibration Error (ECE).
- **User-level:** false alarms per day.

3.5 Main Results

Table 1 summarizes results on the in-house dataset with $\Delta = 6\,\text{h}$. Our method achieves the highest predictive performance and the longest advance warning.

Table 1: Comparison of predictive fall risk models on in-house dataset ($\Delta = 6\,\text{h}$). Best results in bold.

Method	AUROC	AUPRC	Lead-time (min)	False alarms/day
Threshold IMU	0.61	0.24	12.3	5.4
Random Forest	0.68	0.32	18.7	4.8
CNN-LSTM	0.74	0.41	25.6	3.9
Ours (no SSL)	0.80	0.50	33.1	2.7
Ours (full)	**0.86**	**0.59**	**41.5**	**1.9**

Compared to the CNN-LSTM baseline, our model improves AUROC by $+12\%$, extends average lead-time by nearly 16 minutes, and halves daily false alarms.

3.6 Ablation Studies

We perform systematic ablations to isolate contributions:

Effect of multimodal fusion. Removing radar or pressure channels reduces AUROC by 5–8%, confirming complementary benefits.

Effect of self-supervised pretraining. Training from scratch lowers AUPRC by 9%, demonstrating improved representation from masked modeling and contrastive objectives.

Effect of personalization. Without subject-specific adaptation, false alarms increase from 1.9 to 3.2 per day, highlighting the value of lightweight fine-tuning.

Results are summarized in Table 2.

Table 2: Ablation experiments on in-house dataset ($\Delta = 6\,\text{h}$).

Variant	AUROC	AUPRC	False alarms/day
Full model	0.86	0.59	1.9
- w/o radar	0.81	0.49	2.5
- w/o pressure	0.80	0.47	2.8
- w/o SSL	0.77	0.50	2.6
- w/o personalization	0.82	0.55	3.2

3.7 Summary

Overall, the proposed approach consistently outperforms rule-based, classical, and purely supervised deep models. The combination of multimodal sensing, self-supervised pretraining, and personalization yields earlier and more reliable fall risk predictions with fewer false alarms, supporting its practical utility for in-home elder care.

4 Discussion

The results demonstrate that predictive modeling of fall risk is feasible and beneficial when leveraging multimodal sensing, self-supervised representation learning, and personalization. Unlike conventional fall detection systems that react only after an incident, our framework anticipates elevated risk hours in advance, enabling timely interventions. This paradigm shift from detection to prevention has important clinical and social implications.

4.1 Clinical Significance

From a clinical perspective, the ability to forecast falls transforms elder care from reactive to proactive. Caregivers can introduce preventive measures such as mobility aids, physical therapy, or environmental adjustments before accidents occur. The tiered alerting scheme also helps reduce alarm fatigue: by stratifying risk levels, caregivers receive fewer but more meaningful notifications. Case studies revealed that interpretable risk trajectories correlate with clinically recognized precursors, such as gait instability and nocturnal imbalance, thereby enhancing trust and adoption.

4.2 Human–AI Interaction and User Adoption

Beyond technical performance, the effectiveness of predictive fall monitoring systems also depends on how older adults and caregivers perceive and interact with the technology. Prior research in human-computer interaction has shown that users often treat computational systems as social actors [**?**]. Attributing human-like qualities such as reliability, warmth, or authority to algorithmic outputs[5]. In the context of elder care, this means that the presentation of risk alerts - whether they appear as neutral data points, empathetic messages, or authoritative 'advice' - can substantially influence trust, compliance and long-term adoption.

Furthermore, predictive systems operate not only as medical tools but also as social companions[17]. Risk notifications may generate parasocial dynamics: older adults may experience the system as a constant presence of a 'guardian', which can reduce anxiety but may also raise dependency or surveillance concerns[10, **?**]. Therefore, designing the interface and communication strategies around transparency, reassurance, and respect for autonomy is critical.

Ultimately, the diffusion of such systems will hinge on more than just clinical validation. From a communication and marketing perspective, positioning fall prediction as part of a broader "independent living" lifestyle—rather than a stigmatizing medical device—may improve uptake. Culturally adaptive messaging and user-centered onboarding can bridge the gap between technical innovation and everyday acceptance, ensuring that predictive monitoring integrates seamlessly into both the social and domestic lives of aging individuals.

4.3 Future Directions

Future research will focus on three directions. First, integrating additional modalities such as smart flooring or Wi-Fi channel state information may enhance unobtrusive monitoring. Second, reinforcement learning could be explored for adaptive alert thresholds that minimize false positives while ensuring safety. Third, deployment in clinical trials will allow evaluation of downstream outcomes such as reduced hospitalization rates or improved quality of life. Beyond elder care, the methodology may extend to rehabilitation monitoring and chronic disease management where predictive risk stratification is equally valuable.

5 Conclusion

We presented a predictive framework for in-home fall risk assessment that advances beyond traditional detection systems. By combining multimodal sensing, temporal deep learning with self-supervised pretraining, and lightweight personalization, our approach delivers earlier and more reliable forecasts of fall events. Experiments on both naturalistic and public datasets demonstrate improvements in AUROC, lead-time, and false alarm rates compared to rule-based and purely supervised baselines. Importantly, the system provides interpretable risk trajectories that align with clinical observations, fostering trust and practical utility.

This work highlights the potential of moving from *fall detection* to *fall prediction* in smart aging contexts. With continued validation and ethical deployment, predictive fall monitoring may play a crucial role in enabling older adults to live independently and safely at home.

References

[1] Stefano Abbate, Marco Avvenuti, Francesco Bonatesta, Guglielmo Cola, Paolo Corsini, and Alessio Vecchio. A smartphone-based fall detection system. *Pervasive and Mobile Computing*, 8(6):883–899, 2012.

[2] Boris Bačić, Chengwei Feng, and Weihua Li. Jy61 imu sensor external validity: A framework for advanced pedometer algorithm personalisation. *ISBS Proceedings Archive*, 42(1):60, 2024.

[3] Boris Bačić, Claudiu Vasile, Chengwei Feng, and Marian G Ciucă. Towards nation-wide analytical healthcare infrastructures: A privacy-preserving augmented knee rehabilitation case study. *arXiv preprint arXiv:2412.20733*, 2024.

[4] Chengwei Feng, Boris Bačić, and Weihua Li. Sca-lstm: A deep learning approach to golf swing analysis and performance enhancement. In *International Conference on Neural Information Processing*, pages 72–86. Springer, 2024.

[5] Yangfan He, Jianhui Wang, Kun Li, Yijin Wang, Li Sun, Jun Yin, Miao Zhang, and Xueqian Wang. Enhancing intent understanding for ambiguous prompts through human-machine co-adaptation. *arXiv e-prints*, pages arXiv–2501, 2025.

[6] Raul Igual, Carlos Medrano, and Inmaculada Plaza. Challenges, issues and trends in fall detection systems. *Biomedical engineering online*, 12(1):66, 2013.

[7] Feng Jiang, Zongfei Zhang, and Xin Xu. Cmfdnet: Cross-mamba and feature discovery network for polyp segmentation. *arXiv preprint arXiv:2508.17729*, 2025.

[8] Suyuan Li and Xin Song. Future frame prediction network for human fall detection in surveillance videos. *IEEE Sensors Journal*, 23(13):14460–14470, 2023.

[9] Xiang Li and Yikan Wang. Deep learning-enhanced adaptive interface for improved accessibility in e-government platforms. In *2024 6th International Conference on Frontier Technologies of Information and Computer (ICFTIC)*, pages 1692–1695. IEEE, 2024.

[10] Xintao Li, Sibei Liu, Dezhi Yu, Yang Zhang, and Xiaoyu Liu. Predicting 30-day hospital readmission in medicare patients insights from an lstm deep learning model. In *2024 3rd International Conference on Cloud Computing, Big Data Application and Software Engineering (CBASE)*, pages 61–65. IEEE, 2024.

[11] Yuwei Lou, Hao Hu, Shaocong Ma, Zongfei Zhang, Liang Wang, Jidong Ge, and Xianping Tao. Drf: Llm-agent dynamic reputation filtering framework. *arXiv preprint arXiv:2509.05764*, 2025.

[12] Anuradha Singh, Saeed Ur Rehman, Sira Yongchareon, and Peter Han Joo Chong. Sensor technologies for fall detection systems: A review. *IEEE Sensors Journal*, 20(13):6889–6919, 2020.

[13] Lina Tong, Quanjun Song, Yunjian Ge, and Ming Liu. Hmm-based human fall detection and prediction method using tri-axial accelerometer. *IEEE Sensors Journal*, 13(5):1849–1856, 2013.

[14] Xueyi Wang, Joshua Ellul, and George Azzopardi. Elderly fall detection systems: A literature survey. *Frontiers in Robotics and AI*, 7:71, 2020.

[15] Yikan Wang, Chenwei Gong, Qiming Xu, and Yingqiao Zheng. Design of privacy-preserving personalized recommender system based on federated learning. In *2024 8th International Workshop on Materials Engineering and Computer Sciences (IWMECS 2024)*, 2024.

[16] Siye Wu, Lei Fu, Runmian Chang, Yuanzhou Wei, Yeyubei Zhang, Zehan Wang, Lipeng Liu, Haopeng Zhao, and Keqin Li. Warehouse robot task scheduling based on reinforcement learning to maximize operational efficiency. *Authorea Preprints*, 2025.

[17] Peng Zhao, Xiaoyu Liu, Xuqi Su, Di Wu, Zi Li, Kai Kang, Keqin Li, and Armando Zhu. Probabilistic contingent planning based on hierarchical task network for high-quality plans. *Algorithms*, 18(4):214, 2025.

Dynamic Programming and Game-Theoretic Modeling for Resource-Constrained Decision Optimization: A Case Study of the Desert-Crossing Game

Xingyu Zhou[1*], Shuhan Li[2*]

[1]Falcuty of Computer Science and Technology Information, University Malaya, Kuala Lumpur, Malaysia

[2]Falcuty of Science, University Malaya, Kuala Lumpur, Malaysia

Abstract

This paper explores strategic decision-making challenges arising from a desert-crossing simulation game, which serves as a representative model for resource-constrained routing under uncertainty and competition. Multiple scenarios are examined, including single-player settings with known or unknown weather conditions, and multi-player environments with either pre-defined or real-time strategic interactions. Dynamic programming (DP), Dijkstra's algorithm, knapsack problem formulations, and game-theoretic models form the foundation of the analysis. Computational techniques, such as LINGO optimization and Monte Carlo simulation, are applied to obtain and compare solutions. Under fully deterministic weather, a cyclic mining and supply-return route emerges as the optimal strategy. In stochastic environments, simplified heuristics based on expected monetary outcomes are shown to be effective. For multi-agent scenarios, integrating optimal and suboptimal routes reduces resource competition. Dynamic games further benefit from day-by-day route adjustments aimed at minimizing resource consumption while maintaining equitable returns. The study concludes with an evaluation of model performance, identifies current limitations such as validation needs and scenario specificity, and outlines directions for future research, including application to logistics planning and multi-agent system coordination and extensions to more complex network structures and real-world logistics or supply chain problems where resource allocation and pathfinding under uncertainty are crucial.[1]

Keywords—Dynamic programming, Dijkstra's shortest path algorithm, Knapsack problem, Game theory, Nash equilibrium, Stochastic simulation

1 Introduction

1.1 Literature Review

In multistage decision problems, the principle of optimality underpins dynamic programming (DP), first formalized by Bellman[3]. For deterministic routing on nonnegative-weighted graphs, Dijkstra's algorithm efficiently computes shortest paths in $O\big(|E| + |V|\log|V|\big)$ time, and may itself be viewed as a DP instance[6].

Modern routing problems frequently extend these ideas to orienteering variants, where one maximizes collected rewards under time or capacity constraints. For example, Tang et al. embed DP subroutines within reinforcement-learning frameworks to improve orienteering solutions[7]. When uncertainty enters—e.g. stochastic travel times—Monte Carlo Tree Search has been used to solve chance-constrained orienteering problems[5], and rolling-horizon DP with sample-average approximation tackles time-dependent stochastic travel[10].

Meanwhile, noncooperative routing games model competition over shared network resources.[9] Altman and Wynter analyze how self-interested agents reach equilibrium in congested flows via fixed-point iterations akin to DP updates[2]. In discrete route-selection settings, Xu et al. integrate DP-based knapsack solvers into a Nash-equilibrium search, capturing strategic detours under resource contention[8]. While individual techniques are well-established, their integrated application here provides a structured framework for analyzing multi-stage, resource-constrained decisions under varying information structures and competition, offering a testbed for strategies potentially applicable to domains like vehicle routing with refueling or distributed resource gathering.

The "Crossing the Desert" game synthesizes these themes: each stage's cost depends on stochastic weather, inventory constraints resemble knapsack structures, and multiple players compete over limited resources. To our knowledge, no prior work has jointly applied DP, Dijkstra subroutines, and noncooperative game theory in this setting. This paper fills that gap by:

1. Developing single-agent DP models under known and unknown weather scenarios;
2. Embedding these into discrete-time game models for multi-agent competition;
3. Validating strategies via LINGO optimization and C++ Monte Carlo simulation.

1.2 Game Overview

With a map, players use their initial funds to buy a certain amount of water and food (including food and other daily necessities), start from the starting point, and walk in the desert. The goal is to reach the end point within the specified time and keep as much money as possible. You will encounter different

weather conditions along the way, and you can also replenish funds or resources in mines and villages.

1.3 Basic rules of the game

1. Single player game rules

(1) The basic time unit is day. When the game starts, the player is at the starting point, which is recorded as day 0. The player must reach the end point on or before the deadline. After reaching the end point , the game ends for that player.

(2) The weather in all areas of the desert is the same every day, which is one of the three conditions: "clear", "hot" or "sandstorm".

(3) Every day, players can move from one area of the map to another adjacent area, or they can stay where they are. On a sandstorm day, players must stay where they are. (In a map, two areas with a common border are considered adjacent, and two areas with only a common vertex are not considered adjacent.)

(4) Crossing the desert requires two resources: water and food. The minimum unit of measurement is a box. The total weight of water and food a player has each day cannot exceed the upper limit. If the water or food runs out before reaching the end, the game is considered a failure.

(5) The amount of resources consumed by a player staying in one place for one day is the basic consumption, and the amount of resources consumed by a player walking for one day is twice the basic consumption .

(6) When players stay at the mine, they can earn money through mining. The amount of money earned from mining for one day is called basic income. If mining, the amount of resources consumed is three times the basic consumption; if not mining, the amount of resources consumed is the basic consumption. (Mining is not allowed on the day of arrival at the mine, but mining is also possible on sandstorm days.)

(7) On day 0, the player can use the initial funds to purchase water and food at the base price at the starting point. The player can stay at the starting point or return to the starting point, but cannot purchase resources at the starting point multiple times. When the player passes through or stays in the village, he can use the remaining initial funds or the funds obtained from mining to purchase water and food at any time. The price per box is twice the base price. After reaching the end point, the player can return the remaining resources at half the base price.

2. Multiplayer Game Rules

(1) If any of the players walk from area A to area B on a certain day, the amount of resources consumed by any of them is k times the basic consumption.

(2) If any of the players mine in the same mine on a certain day, the amount of resources consumed by any of them is 3 times the basic consumption, and the funds that each player can obtain through mining in a day are the basic income .

(3) If any of the players purchase resources in the same village on a certain day, the price of each box will be 4 times the base price.

(4) In other cases, the amount of resources consumed and the price of resources are the same as those in a single-player game.

1.4 Problem

Question 1: There is only one player, and the weather conditions are known in advance every day during the entire game period. Give the optimal strategy for the player under normal circumstances. Solve the "First Level" and "Second Level" in the attachment. The remaining funds (remaining water, remaining food) refers to the funds (water, food) after all the resources required for the day have been consumed. If there is any purchase behavior on the same day, it refers to the funds (water, food) after the purchase is completed.

Question 2: There is only one player, and the player only knows the weather conditions of the day. He can decide the action plan for the day based on this. Give the best strategy for the player under normal circumstances, and discuss the "third level" and "fourth level" in the attachment in detail.

Question 3: There are n players who have the same initial capital and start from the starting point at the same time.

(1) Assume that the weather conditions for each day of the game are known in advance. Each player's action plan must be determined on day 0 and cannot be changed thereafter . Give the strategies that players should take in general, and discuss the "fifth level" in detail.

(2) Assume that all players only know the weather conditions for the day. Starting from the first day , each player knows the action plans of other players and the amount of resources left after completing the action for the day, and then determines their own action plan for the next day. Give the strategies that players should adopt in general, and discuss the "Sixth Level" in detail.

2 Model Development and Solution Approaches

2.1 Explanation of symbols

2.2 Analysis of Question 1

According to the rules of the game, the game is roughly calculated and analyzed, and the optimal strategy is determined according to different situations. Thus, multiple plans are determined, such as going directly from the starting point to the end point, from the starting point to the mine or village, and then back to the end point. Then one or more plans are selected according to the specific situation of different levels, and the remaining funds of the player after reaching the end point under different plans are calculated, and the optimal plan is obtained by comparison.

2.3 Analysis of Question 2

Since players can only know the weather conditions of the day, in general, they can simply estimate the local weather condi-

Table 1: List of Symbols and Definitions

Symbol	Meaning
R_w	Total expenditure amount
R_{w_1}	Expenditure on goods purchased in the initial village
R_{w_2r}	Expenditure on goods purchased in the village during the r-th loop
R_{w_2}	Total expenditure on goods purchased during transit
W_0	Initial capital
d_i	Unit price of item i
k_i	Quantity of item i purchased
P	Base income
t_j	Number of days required in stage j
t_{jq}	Number of sunny days in stage j
t_{jg}	Number of hot days in stage j
t_{js}	Number of sandstorm days in stage j
X_{ij}	Amount of item i consumed in stage j
b_{iq}	Consumption of item i on a sunny day
b_{ig}	Consumption of item i on a hot day
b_{is}	Consumption of item i on a sandstorm day
c_i	Weight of item i
P_{ij}	Plan option j under situation i
M_i	Maximum number of boxes for item i
$f[i][v]$	DP state variable in the knapsack problem
R'_w	Revenue from residual goods at the destination
R_{mn}	Position of person m after the n selection
W_{mn}	Remaining capital of person m after the n selection

tions based on the number of days, choose the specific direction of movement for the day, and buy more resources based on the shortest path from the starting point to the end point to provide more options for the player's movement. For the third level, two solutions can be obtained through analysis. Under the condition of randomly generating enough weather patterns, the expected value of the funds left over from the two solutions after n games with different weather conditions is calculated to determine the optimal solution. For the fourth level, based on the basic model of problem one, four solutions for problem four are obtained, and then the four solutions are simplified using the characteristics of the graph and known conditions. Using similar ideas to the third level, the expected value of each solution under enough weather conditions is used to measure the advantages and disadvantages of the solution.

2.4 Analysis of Question 3

In the first question, since there are multiple players participating in the game at the same time, the optimal route when there is only one player is no longer the optimal route. Therefore, the general strategy of the player is to randomly choose one of the optimal and suboptimal routes as the action plan for this game on day 0. When solving the "fifth level", first consider the shortest route, three days, four days, and five days. Considering that the two players are in a competitive relationship, both players consider routes with less consumption or more benefits, so only consider both choosing a three-day route, both choosing a four-day route, or one of them choosing a three-day route and one choosing a four-day route, or both choosing a five-day mining route. Introducing random numbers and using C++ for simulation, we obtained the amount of funds left when the player reaches the end in all the cases we listed.

In the second question, because after the end of each day's game, players will determine tomorrow's action plan based on their own funds and resources, and know other players' funds and resources to infer other players' action plans. In this case, players need to choose the movement plan that consumes the least funds and resources for the next day. When conducting a specific analysis and discussion of the "Sixth Level", 0-1 variables are introduced to control the weather. Let the unknown location of the mth person after the nth iteration be , combined with the map of the "Sixth Level", dynamic data prediction for players is carried out, and the restrictions on players' mining, path selection, and whether to enter the village are discussed.

3 Scenario 1: Single Player with Deterministic Weather

3.1 Determination of the Optimal Strategy

The overall decision direction of the player is: staying in areas other than the mine will consume player resources and bring no benefits, so it is not considered that the player will stay in other areas in sunny or hot weather, and the player will not return to the starting point after departure.

If you decide to go mining, the player's general decision direction is: while ensuring that water and food are consumed roughly at the same time, try to buy more food at the starting point, buy more water and less food in the village; after reaching the end, try to avoid having surplus resources.

Case 1: At the starting point, decide whether to travel to the mine to extract resources based on the mine's basic daily income.

Decision 1: If the basic daily income of the mine is far less than the funds required by the player to go to the mine, the optimal strategy is to purchase only the resources that the player needs from the starting point to the end point at the starting point, without going to the mine, and go directly to the end point.

The decision to undertake a mining detour is formulated as a cost-benefit analysis. Let ΔC_{mine} represent the **additional cost** (in monetary units) of traveling to the mine, mining for m days, and then proceeding to the final destination, as compared to taking the most direct path to the destination. This cost is derived from the weather-dependent consumption of resources (water and food) along the longer path and during mining activities.

The strategy of choosing to mine is optimal if the total revenue generated exceeds this additional cost. This is formalized by the following inequality:

$$p \cdot m > \Delta C_{\text{mine}} \tag{1}$$

where:

- p is the base daily income from mining (a constant given in the game rules).
- m is the number of days the player chooses to spend mining ($m \geq 1$).
- ΔC_{mine} is the net additional cost of the mining detour.

Decision 2: The player should opt for the mining strategy if and only if there exists an integer $m \geq 1$ such that inequality (1) holds and the player's initial resources and the time constraint allow for this detour.

Case 2: At the starting point, calculate the funds needed for the player to go to the village and to the mine based on the weather. Based on the comparison of the funds, decide whether to go to the mine or the village first.

Decision 3: If the funds required to reach the mine are small, then go to the mine first. If the remaining resources after reaching the mine are sufficient to support the player to reach the village to replenish resources after digging for at least one day, then the best decision is to reach the mine first.

Decision 4: If the funds required to reach the village are less, then go to the village first.

Case 3: The player makes the best decision based on the actual situation in the mine or village. And the player can go back and forth between the mine and the village.

Situation 3-1: In the mine, on the premise of ensuring that the player can reach the end point from the mine, the player decides whether to go to the village to replenish resources based on the remaining resources, funds and weather conditions of the remaining days.

Decision 5: If the remaining resources, funds, and days are sufficient to support the player to go to the village to replenish resources and the income from returning to the mine to mine is greater than the player's capital consumption to replenish resources, then the optimal decision is to go to the village to replenish resources and return to the mine to mine.

Decision 6: If the player goes to the village to replenish resources at the end of the remaining days, and the income from returning to the mine to mine is less than the cost of replenishing resources, then the optimal strategy is for the player not to replenish resources and return directly to the end point.

Situation 3-2: In the village, decide whether to go back to the mine to mine based on the remaining resources, funds and number of days.

Decision 7: If the player's mining income is greater than the capital expenditure, the optimal strategy is for the player to travel to the mine to extract resources after replenishing resources in the village.

Decision 8: If the player's mining income is less than the capital expenditure, the optimal strategy is for the player not to mine and return to the end point.

3.2 Establishment of general model

3.2.1 Model preparation - establishment of description matrix

Based on the map that the player gets when starting the game, use a 0-1 matrix to describe whether area A and area B are adjacent.

$$\alpha_{ij} \begin{cases} 1 & \text{if } v_i \text{ and } v_j \text{ share a common boundary} \\ 0 & \text{if } v_i \text{ and } v_j \text{ do not share a common boundary} \end{cases}$$

The matrix A_0 can be obtained as follows:

$$A_0 = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}$$

3.2.2 Model from the starting point to the end point directly

If the player does not consider going to the mine to mine , then he can directly consider the shortest path from the starting point to the end point . The shortest path is certain, and the money to be spent on buying water and food is also certain , so after determining the shortest path , the player's money at the end point can also be determined accordingly. (In the case where the shortest path passes through a mine or a village, we will divide it into the following scheme for discussion).

Using Dijkstra's shortest path graph theory algorithm, we can get the following model:

Model 1 (All-Pairs Shortest Path): Let $A^{(k)}(i, j)$ represent the length of the shortest path from node i to node j where

only nodes $1, 2, \ldots, k$ are allowed as intermediate nodes.

$$A^{(k)}(i,j) = \min\left(A^{(k-1)}(i,j),\ A^{(k-1)}(i,k) + A^{(k-1)}(k,j)\right) \tag{2}$$

for $k = 1$ to n, where n is the total number of nodes (areas) on the map.

The initial state is:

$$A^{(0)}(i,j) = \begin{cases} 0 & \text{if } i = j \\ w(i,j) & \text{if } i \neq j \text{ and an edge exists with weight } \\ \infty & \text{otherwise (no direct connection)} \end{cases}$$

3.2.3 Model from the starting point to the end point via a mine or village

1. Preliminary classification of models

Based on the above optimal strategy, we first preliminarily determined three options, namely, going to the mine first and then to the village, going to the village first and then to the mine, and going back and forth between the mine and the village.

A simple analysis shows that the goal of the game is to reach the end within the specified time and keep as much money as possible. Considering that the basic income of the mine is relatively rich, walking or staying in areas other than the mine will only consume resources, so we consider excluding the necessary time to move between the starting point, the mine and the village, and the end point, so that the player can only stay in the mine as much as possible. We establish mathematical models corresponding to these three solutions respectively. When solving the levels, we add, reduce or exchange the intermediate links according to the specific situation, and introduce known information to simplify the model of each level.
The first level is shown in the Fig.1.

2. Model establishment for the scenario where the player goes to the village first and then to the mine

Model 2: The player's walking diagram is as follows: The objective function is to maximize the remaining funds at the endpoint:

$$max\ w = w_0 + p(t_0 - 1) + R'_w - R_w \tag{3}$$

where w_0 is the initial funds, P is the player's basic income at the mine, R'_w is the funds obtained by returning resources at the endpoint, and R_w is the total funds spent on purchasing resources.

$$R_w = R_{w_1} + R_{w_2} \tag{4}$$

Here, $R_{w_1} = \sum_{i=1}^{2} d_i k_i$ represents the funds spent on purchasing k_1 boxes of water and k_2 boxes of food at the starting point, and $R_{w_2} = 2\sum_{i=1}^{2} d_i k_{i+2}$ represents the funds spent on purchasing k_3 boxes of water and k_4 boxes of food in the village. d_i denotes the unit price. The total funds spent must not exceed the initial funds.

Using the knapsack problem in dynamic programming, determine the ratio of water and food the player should purchase

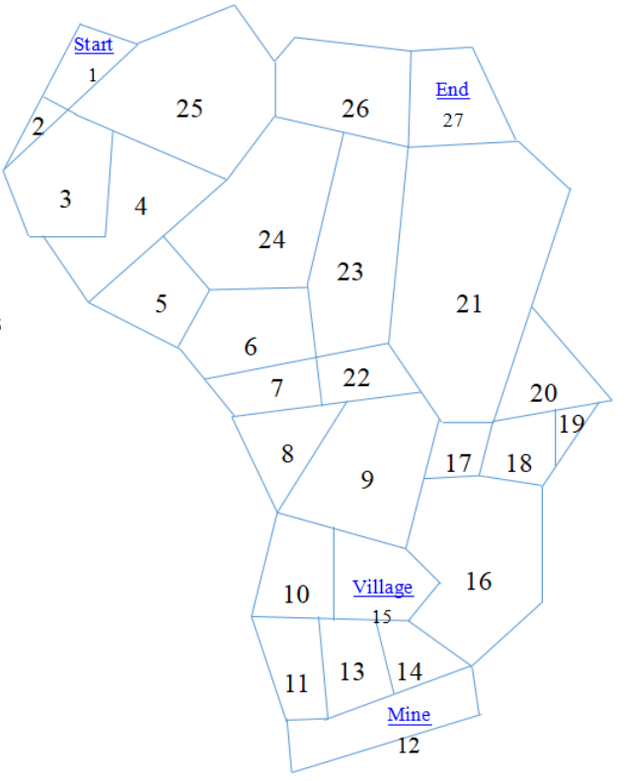

Figure 1: "Level 1" map

at the starting point:

$$f[i][v] = \max\left\{ f[i-1][v - k_i C_i] + k_i w_i \;\middle|\; 0 \le k_i \le M_i \right\} \tag{5}$$

$$\text{where } i \in \{1,2\}, \quad m \in \{1,2,3,4\}$$

The player's game duration is 30 days:

$$t_j = t_{jq} + t_{jg} + t_{js}, \quad j \in \{1,2,3,4\}, \quad t_j \le 30 \tag{6}$$

The resource consumption boxes without mining are:

$$X_{ij} = 2(b_{iq}t_{jq} + b_{ig}t_{jg}) + b_{is}t_{js}, \quad j \in \{1,2,3\} \tag{7}$$

The resource consumption boxes for mining control are:

$$X_{ij} = 3(b_{qj}t_{jj} + b_{gj}t_{fg} + b_{hj}t_{hs}), \quad j = 4 \tag{8}$$

At the endpoint, the player can return remaining resources at half price, with the resulting income:

$$R'_w = \frac{1}{2}\sum_{i=1}^{2} d_i \left(k_i + k_{i+2} - \sum_{j=1}^{4} X_{ij} C_j \right) \tag{9}$$

The total weight of resources owned by the player must not exceed the load capacity limit:

$$\sum_{i=1}^{2} C_i(k_i + k_{i+2}) - \sum_{j=1}^{4} X_{ij} C_j \le M \tag{10}$$

For clarity, the core step in many shortest path algorithms involves comparing paths through intermediate nodes. A standard formulation for computing the shortest path between all pairs of nodes (like the Floyd-Warshall algorithm) is: This formulation iteratively improves the shortest path estimate between nodes *i* and *j* by considering paths through an intermediate node *k*. Additionally, although this scenario does not follow the shortest path, imposing a shortest path constraint restricts unnecessary player movements and reduces resource waste:

Let $A^{(k)}(i,j)$ represent the length of the shortest path from node i to node j where only nodes $1, 2, \ldots, k$ are allowed as intermediate nodes.

$$A^{(k)}(i,j) = \min\left(A^{(k-1)}(i,j),\ A^{(k-1)}(i,k) + A^{(k-1)}(k,j)\right) \tag{11}$$

for $k = 1$ to n, where n is the total number of nodes (areas) on the map.

The initial state is:

$$A^{(0)}(i,j) = \begin{cases} 0 & \text{if } i = j \\ w(i,j) & \text{if } i \neq j \text{ and an edge exists with weight } w(i,j) \\ \infty & \text{otherwise (no direct connection)} \end{cases}$$

3. Modeling of the scenario where the player goes to the mine first and then to the village

The player's walking diagram is as follows:

Model 3: Similarly, the following model can be established:

$$\max w = w_0 + p(t_4 - 1) + R'_w - R_w$$

$$R_w = \sum_{i=1}^{2} d_i k_i + 2\sum_{i=1}^{2} d_i k_{i+2} \leq w_0$$

$$f[i][v] = \max\{f[i-1][v - k_i C_i] + k_i w_i, \geq 0\}$$

$$i \in \{1,2\}, \quad m \in \{1,2,3,4\}$$

$$0 \leq k_i \leq M, \quad k_i - \sum_{j=1}^{m} X_{ij}$$

$$R_{w1} \leq w_0$$

$$R_{w2} \leq w_0 - R_{w1} + p(t_4 - 1)$$

$$t_j = t_{jq} + t_{jg} + t_{js}, \quad j \in \{1,2,3,4\}$$

$$X_{ij} = 2\left(b_{qf_{jq}} + b_{gf_{jg}}\right) + b_{gf_{jg}}, \quad j \in \{1,2,3\}$$

$$X_{ij} = 3\left(b_{qf_{jq}} + b_{gf_{jg}}\right) + b_{gf_{jg}}, \quad j = 4$$

$$R'_w = \frac{1}{2}\sum_{i=1}^{2} d_i \left(k_i + k_{i+2} - \sum_{j=1}^{4} X_{ij}\right)$$

$$\sum_{i=1}^{2} C_i(k_i + k_{i+2}) - \sum_{i=1}^{2}\sum_{j=1}^{4} X_{ij} C_i \leq M$$

$$A^{(k)}(i,j) = \min\left(A^{(k-1)}(i,j),\ A^{(k-1)}(i,k) + A^{(k-1)}(k,j)\right)$$

The differences from **Model 3** compared to **Model 2** are:

1. Constraint 1:

$$R_{v1} \leq w_0$$

This means the funds spent by the player at the starting point cannot exceed the initial funds.

2. Constraint 2:

$$R_{v2} \leq w_0 - R_{w1} + p(t_4 - 1)$$

This indicates that the funds spent by the player in the village must not exceed the sum of the remaining funds after starting point purchases and the income earned from mining.

3. Dynamic Programming Adjustment:

$$f[j][v] = \max\{f[i-1][v - k_i C_t] + k_i w_t, \quad 0 \leq k_i \leq M_t, \}$$

This formula specifies that the resources purchased at the village must support the player's journey from the starting point to the mine, mining activities, and resource replenishment from the mine to the village.

4. The establishment of a model for players to travel back and forth between the mine and the village

The player's walking diagram is as follows, in which players can travel back and forth between the mine and the village based on their existing funds, resources, remaining days and weather, without losing funds.

On the basis of **Model 2** and **Model 3**, we introduce the parameter r to represent the number of times the player goes back and forth between the mine and the village, and establish the mathematical model of this scheme:

Model 4:

$$\max w = w_0 + p\sum_{i=1}^{2}(t_{4r} - 1) + R_w - R_w$$

$$R_w = R_{w_1} + R_{w_2}$$

$$R_{v_1} \leq w_0$$

$$R_{W_3} = 2\sum_{i=1}^{2} d_i k_{(i+2)} r$$

$$R_{v_2} r \leq w_0 - R_{v_1} + p(t_{4r} - 1) - 2\sum_{i=1}^{2} d_i k_{(i+2)(r-1)}$$

$$f[i][v] = \max\{f[i-1][v - k_i C_i] + k_i w_i, \quad 0 \leq k_i \leq M_i\}$$

$$i \in \{1,2\}, \quad m \in \{1,2,3,4\}$$

$$t_j = t_{jq} + t_{jg} + t_{js}, \quad j \in \{1,2,3,4\}$$

$$t = t_1 + t_2 + r(t_3 + t_4) \leq 30$$

$$X_{ij} = 2(b_{iq} t_{jq} + b_{ig} t_{jg}) + b_{is} t_{js}, \quad j \in \{1,2,3\}$$

$$X_{ij} = 3(b_{iq} t_{jq} + b_{ig} t_{jg} + b_{is} t_{js}), \quad j = 4$$

$$k_{3n} - \sum_{i=2}^{3} X_{3i} \geq 0, \quad k_{4n} - \sum_{i=2}^{3} X_{4i} \geq 0 \quad (r = n)$$

$$R'_w = \frac{1}{2}\sum_{i=1}^{2}\sum_{r=1}^{n} d_i \left(k_{ir} + k_{(i+2)r} - \sum_{j=1}^{4} X_{ij}\right)$$

$$\sum_{i=1}^{2} C_i(k_i - X_{i1}) + \sum_{i=1}^{2} X_{i1}k_{(i+2)r} - r\sum_{i=1}^{2}\sum_{\substack{j=2\\ j\neq 3}}^{4} X_{ij}C_i \leq M$$

$$A^{(k)}(i,j) = \min\left(A^{(k-1)}(i,j),\ A^{(k-1)}(i,k) + A^{(k-1)}(k,j)\right)$$

Where:

1. $R_{T_2} = 2\sum_{i=1}^{2} d_i k_{(i+2)} r$ represents the money spent on purchasing resources in the village r times;

2. $R_{v2}r \leq w_0 - R_{w1} + p(t_{4r} - 1) - 2\sum_{i=1}^{2} d_i k_{(i+2)(r-1)}$ (when $r = 0, R_{v2}r = R_{w1}$) where means that the funds spent on purchasing resources in the village each time cannot exceed the sum of the remaining funds after purchasing resources at the starting point and the income from the mine this time minus the funds spent on the last purchase in the village.

3. $t = t_1 + t_2 + r(t_3 + t_4) \leq 30$ which means the total number of game days cannot exceed 30 days;

4. $k_{3n} - \sum_{i=2}^{3} X_{3i} \geq 0, \quad k_{4n} - \sum_{i=2}^{3} X_{4i} \geq 0 \quad (r = n)$ which means that the player had enough water and food when he returned to the end point from the village or mine for the last time.

3.3 Model building and solution of the "First level"

3.3.1 Description Matrix for details)

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

3.3.2 Model solution and result display

1. Solution steps

Use LINGO software to solve the model:

Step 1: Use Dijstra's shortest path algorithm to find the five shortest paths from the starting point to the village or mine, from the end point to the village or mine, and from the starting point to the end point.

Step 2: Substitute the shortest path into the four basic models to calculate the optimal solution among the four solutions. When solving, the description matrix needs to be flipped and solved twice.

Step 3: Sort out the optimal route according to the output of Model 4 of the optimal solution

2. Results display

The following table can be obtained after sorting.

The player returned to the starting point on the 23rd day. Draw the route on the map of "Level 1" as shown below Fig. 2. As shown below Table. 2.

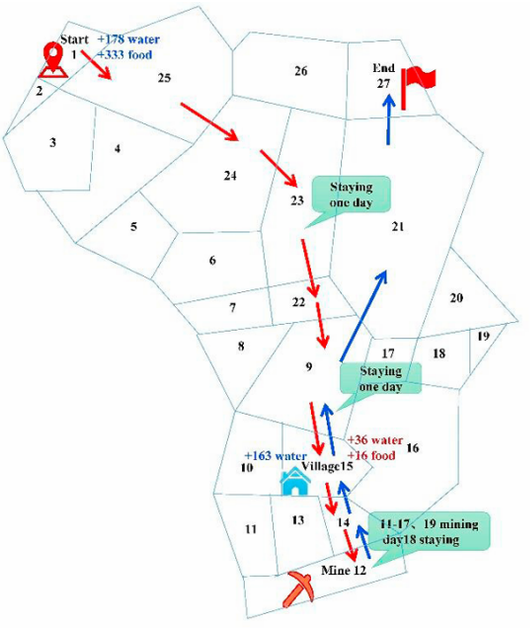

Figure 2: The best route map for the "Level 1"

3.3.3 Model building and solution of the "Second level"

1. Description matrix:

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

2. Model solution steps and model display

The Second level is shown in the Fig. 3.

The solution steps are similar to those of the first level, except that the description matrix and the total number of map areas are replaced. We also use LINGO to solve the model according to the route Table. 3.

Draw the route on the map of "Level 2" as shown below Fig. 4.

The line in the picture is the walking path, and the area is the area where the player needs to stay.

4 Scenario 2: Single Player with Stochastic Weather

4.1 Model establishment and solution of problem 2: Optimal Strategy

Since the player can only know the weather conditions on the current day, the weather on the day the player plays the game and before that is known.

1. If there are few game days, players cannot make a simple prediction of the weather in that place. Then players can only

Table 2: ”First Level” Results Table

Day	Location	Remaining Funds	Remaining Water
0	1	5800	180
1	25	5800	164
2	24	5800	148
⋮	⋮	⋮	⋮
21	9	10430	26
22	21	10430	16
23	27	10430	0
24			
25			
26			
27			
28			
29			
30			

Table 3: ”Second Level” Results Table

Day	Location	Remaining Funds	Remaining Water
0	1	6475	247
1	2	6475	231
2	3	6475	215
⋮	⋮	⋮	⋮
28	55	14775	32
29	63	12365	16
30	64	12365	0

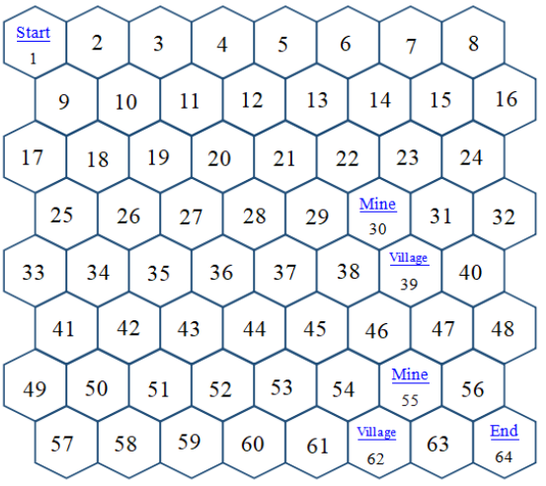

Figure 3: ”Level 2” map

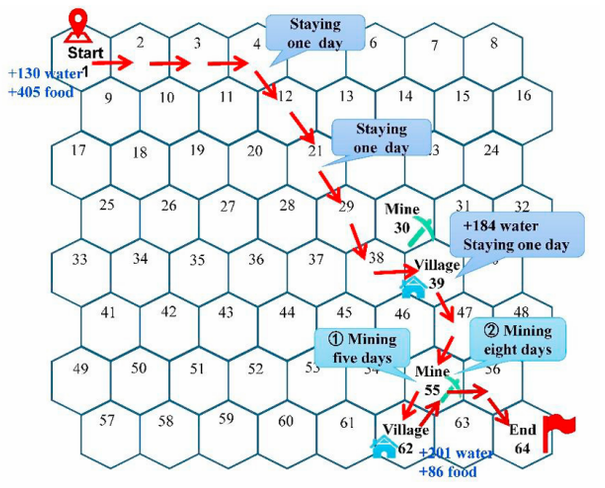

Figure 4: The best route map for the ”Level 2”

make decisions based on the conditions of the day. If it is sunny, choose a moving direction that is close to the end point and the village or mine. If it is hot weather, players choose to move in the shortest direction from their current location to the end point.

2. If there are many game days, move according to the strategy in the first few days of the game. When the player has played the game for a certain number of days and can make a rough estimate of the weather in that place, he can decide whether to go to the mine or village based on his own resources and funds, or directly return to the end point by the shortest path from his current location.

3. If the player has reached the village or mine, he can make the best decision based on situation 3 in question 1.

4. Therefore, at the starting point, players need to buy an appropriate amount of resources that are more than what is needed for the shortest path based on the length of the map.

4.2 Establishment and solution of the “third level” model

1. Graphic analysis and solution determination

In question one, we gave four basic solutions: going directly from the starting point to the end point, from the starting point to the mine or village and then to the end point. By analyzing the image characteristics of the third level below, we found that there is no village on the Fig. 5.

Therefore, we simplify the four basic solutions into two solutions suitable for this level according to the specific situation:

Solution 1: Take the shortest path from the starting point to the end point;

Option 2: Take the shortest path from the starting point to the mine for mining and then take the shortest path back to the end point.

2. Establishment of the corresponding solution model

It is known that there will be no sandstorm weather within 10 days, so we define t_1 as days of high-temperature weather and t_2 as days of clear weather:

$$t_1 + t_2 = 10 \tag{12}$$

Since the graph is relatively simple, we directly obtain the shortest path from the starting point to the endpoint (3 days) and from the mine back to the endpoint (2 days). Assume the

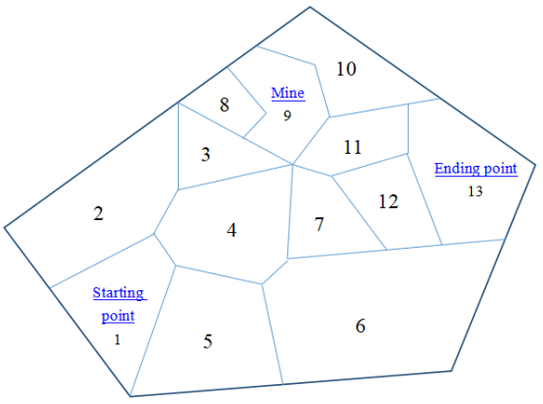

Figure 5: "Level 3" map

first 3 days consist of t_3 days of high-temperature weather and t_4 days of sunny weather:

$$t_3 + t_4 = 3 \tag{13}$$

Model for Scheme 1:

$$w_1 = w_0 - 2\sum_{i=1}^{2}(d_i b_{ig} t_3 + d_i b_{ig} t_4) \tag{14}$$

$$t_3 + t_4 = 3$$

Assume mining days include t_5 days of high temperatures and t_6 days of clear skies, ensuring the player can return smoothly from the mine to the endpoint:

$$t_5 + t_6 \leq 5 \tag{15}$$

Model for Scheme 2:

$$w_2 = w_0 - 2\sum_{i=1}^{2}[d_i b_{ig}(t_1 - t_5) - d_i b_{ig}(t_2 - t_6)] \tag{16}$$

$$t_1 + t_2 = 10$$
$$t_5 + t_6 \leq 5$$

3. Comparison of the two solutions determined

Since the player only knows the current day's weather condition, the specific weather pattern throughout the game cannot be predetermined. Thus, direct comparison of w_1 and w_2 is infeasible. We randomly generate n weather patterns, including all high-temperature days, all clear days, first 5 days as high-temperature followed by 5 clear days, etc.:

$$t_i \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}, \quad (i \in \{1, 2, 3, 4, 5, 6\})$$

where t_i and t_{i+1} satisfy Equations 12 and 13.

For each weather pattern, the player follows both **Scheme 1** and **Scheme 2** in the game. The expected remaining funds for each scheme after n simulations are calculated as:

$$\mathbb{E}[w_1] = \frac{\sum_{i=1}^{n} w_{1i}}{n}, \quad \mathbb{E}[w_2] = \frac{\sum_{i=1}^{n} w_{2i}}{n}$$

When n is sufficiently large, the relative magnitudes of these expectations reflect the superiority of the two schemes.

4. Model solution and results

Based on the model of the first question, a random factor is introduced to describe the weather changes, and the description matrix and the total number of map areas are also changed. The map description matrix of the "third level" is:

$$\begin{bmatrix} 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$

Scheme 1: Mean remaining funds $\mu_1 \approx 9925$, Standard Deviation $\sigma_1 \approx 25$;
Scheme 2: Mean remaining funds $\mu_2 \approx 9139.25$, Standard Deviation $\sigma_2 \approx 20$;
Since $\mu_1 > \mu_2$, and considering the variances, Scheme 1 appears statistically better under the simulated conditions.
After 50 random operations, we get:

$$\overline{W}_1 \approx 9925, \quad \overline{W}_2 \approx 9139.25, \quad \overline{W}_1 > \overline{W}_2,$$

so plan 1 is better.

4.3 Establishment and analysis of the "fourth level" model

1. Determination of the plan

Since the map of the fourth level has both villages and mines, it conforms to the four basic solutions we determined in question one. However, since we only know the weather on that day in this question, in order to introduce the parameter t, we adjust the four basic solutions and obtain four solutions that are more in line with this question.

Plan 1: Take the shortest path from the starting point to the end point;

Plan 2: Start from the starting point, go to the mine and then go to the end point;

Plan 3: Start from the starting point to the mine and then go to the end point;

Plan 4: From the starting point to the mine or village, travel back and forth between the mine and the village according to actual conditions, and finally go from the village or mine to the end point.

2. Establishment of the corresponding solution model

Under the premise of the four basic schemes in Problem 1, introduce the parameter t.

a. Assume there are t_1 days of high-temperature weather, t_2 days of clear/cloudy weather, and t_3 days of sandstorm weather.

Model for Scheme 1:

$$w_1 = w_0 - \sum_{i=1}^{2} d_i k_i$$

$$\begin{cases} \sum_{m=1}^{2} t_m = 8 \\ k_i - 2\left(b_{iq}t_1 + b_{ig}t_2\right) - b_{is}t_3 \geq 0, \quad i \in \{1,2\}, \quad 0 \leq k_i \leq M_i \end{cases}$$

b. Assume mining includes t_4 days of high-temperature weather, t_5 days of clear/cloudy weather, and t_6 days of sandstorm weather.

Model for Scheme 2:

$$w_2 = w_0 - \sum_{i=1}^{2} d_i k_i + p(t_4 + t_5 + t_6 - 1)$$

$$\begin{cases} \sum_{m=1}^{3} t_m \leq 30 \\ k_i = 2\left[b_{iq}(t_1 - t_4) - b_{ig}(t_2 - t_5)\right] - b_{is}(t_3 - t_6) \geq 0, \quad i \in \{1,2\} \end{cases}$$

c. Assume the player purchases water k_3 and food k_4 in the village.

Model for Scheme 3:

$$w_3 = w_0 - \sum_{i=1}^{2} d_i k_i - 2\sum_{i=1}^{2} d_i k_{i+2} + \frac{1}{2}\sum_{i=1}^{2}\left(k_i + k_{i+2} - \sum_{j=1}^{3} X_{ij}\right)$$

$$\begin{cases} \sum_{m=1}^{3} t_m \leq 30 \\ k_1 + k_3 - 2\left(b_{1q}t_1 + b_{1g}t_2\right) - b_{1s}t_3 \geq 0 \\ k_2 + k_4 - 2\left(b_{2q}t_2 + b_{2g}t_2\right) - b_{2s}t_3 \geq 0, \quad 0 \leq k_i \leq M \\ X_{ij} = 2\left(b_{iq}t_{jq} + b_{ig}t_{gj}\right) + b_{is}t_{js}, \quad j \in \{1,2,3\} \\ \sum_{i=1}^{2} C_i\left[k_i + k_{i+2} - 2\left(b_{iq}t_1 + b_{ig}t_2\right) - b_{is}t_3\right] \leq M \end{cases}$$

d. Assume mining includes t_4 days of clear weather, t_5 days of high-temperature weather, t_6 days of sandstorm weather, with $1 \leq r \leq n$.

Model for Scheme 4:

$$w_4 = w_0 + p\sum_{r=1}^{n}\left(t_{4r} + t_{5r} + t_{6r} - 1\right) - \sum_{i=1}^{n} d_i k_i$$

$$\begin{cases} \sum_{m=1}^{3} t_m \leq 30 \\ k_1 + k_{3r} - 2\left(b_{1q}t_{1(r-1)} + b_{1g}t_{2(r-1)}\right) - b_{1s}t_{s(r-1)} \geq 0 \\ k_2 + k_{4r} - 2\left(b_{2q}t_{2(r-1)} + b_{2g}t_{2(r-1)}\right) - b_{2s}t_{s(r-1)} \geq 0 \\ X_{ij} = 2\left(b_{iq}t_{jq} + b_{ig}t_{js}\right) + b_{is}t_{js}, \quad j \in \{1,2,3\} \\ X_{ij} = 3\left(b_{iq}t_{jq} + b_{ig}t_{js} + b_{is}t_{js}\right), \quad j \in \{4\} \end{cases}$$

3. Comparison of four solutions

So we randomly generate n weather patterns, including all high-temperature days, all clear days, etc.:

$$t_i \in \{0,1,2,3,4,5,6,7,8,9,10\}, \quad (i \in \{1,2,3,4,5,6\})$$

Since Level 4 rarely experiences sandstorms within 30 days, we impose:

$$t_1 = t_2 = 10t_3$$

For each weather pattern, the player follows all four schemes in the game. The expected remaining funds for each scheme after n simulations $(\mathbb{E}[w_1], \mathbb{E}[w_2], \mathbb{E}[w_3], \mathbb{E}[w_4])$ are calculated. When n is sufficiently large, the relative magnitudes of these expectations reflect the superiority of the four schemes.

4. Specific analysis of the combination of model and map

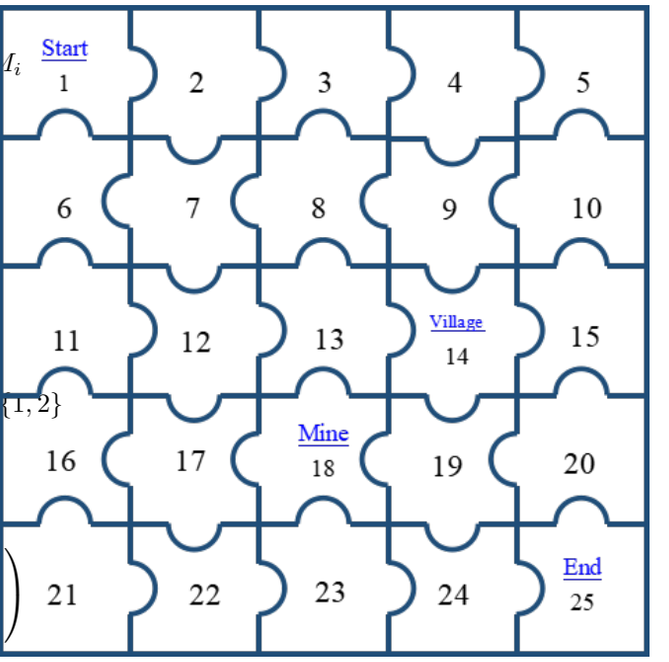

Figure 6: "Level 4" Map

(1) Since the village and the mine are both on the shortest path from the player's starting point to the midpoint, and the price of materials in the village is twice that of the starting point, we do not consider option 3, that is, we do not consider going directly back to the end point after shopping in the village from the starting point. At the same time, the income from the mine is relatively high, so we do not consider option 1, which is to take the shortest path directly to the end point. Draw the route on the map of "Level 4" as shown below **Fig. 6**

(2) Analysis shows that the shortest paths from the starting point to the mine and the village are the same. Due to the uncertainty of the weather, the player cannot decide the quantity of resources to purchase if he goes to the village first, so the player considers going to the mine first to mine.

(3) In the mine, the player is guaranteed to reach the end point by calculating the remaining resources, funds and weather conditions of the remaining days.

a. If the remaining resources, funds, and days are sufficient to support the player to go to the village to replenish resources and the income from returning to the mine to mine is greater than the player's capital consumption to replenish resources, then the optimal decision is to go to the village to replenish resources and return to the mine to mine.

b. If the player goes to the village to replenish resources at the end of the remaining days, and the income from returning to the mine to mine is less than the cost of replenishing resources, the optimal strategy is for the player not to replenish resources and return directly to the end point.

5. Model solution and results

The description matrix of the "fourth level" is:

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

From the above analysis, we know that plan 4 is better. We use LINGO to calculate the model, and the results are consistent with our analysis. The specific plan is: travel to the mine to extract resources first, then go to the village to replenish resources, then travel to the mine to extract resources, and finally return to the end. When the player returns to the end, the remaining funds are 13625 yuan.

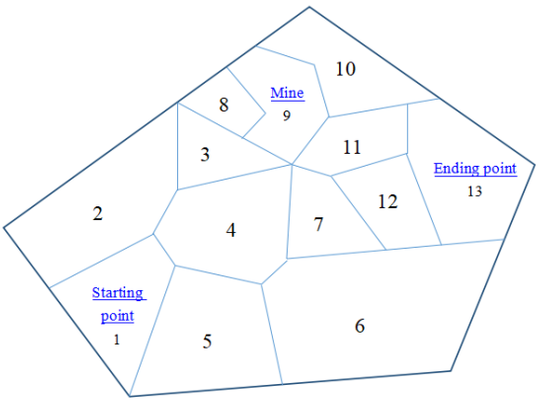

Figure 7: "Level 5" map

5 Scenario 3: Multi-player Competitive Setting

5.1 If the weather is known in advance and the plan is not changed after the starting point is determined

5.1.1 Strategies that players should generally adopt

Since multiple players participate in the game simultaneously, the following rules apply:

1. If k $(2 \leq k \leq n)$ players move from area A to B on the same day, the resource consumption becomes $2k$ times of the base consumption.

2. If k $(2 \leq k \leq n)$ players mine in the same mine on the same day, each player's resource consumption is 3 times of the base consumption, and their daily mining income is reduced to $\frac{1}{k}$ times of the base income.

3. If k $(2 \leq k \leq n)$ players purchase resources in the same village on the same day, the price per box increases to 4 times of the base price.

This will cause the optimal route when there is only one player to no longer be the optimal route, so players should consider the suboptimal route in the range of options when choosing a plan on day 0.

Therefore, the general strategy of the player is to randomly choose one of the optimal route and the suboptimal route as the action plan for this game on day 0. This resembles a mixed strategy in game theory, where players randomize over pure strategies (routes) to maximize their expected payoff given the potential actions of others, aiming for a Nash equilibrium where no player can benefit by unilaterally changing their strategy.

Without considering other reasons for longer routes, the probability of players choosing a longer route is small, and choosing a longer route does not effectively reduce the probability of meeting other players, so the resources consumed by players with a high probability of choosing a longer route will increase.Draw the route on the map of "Level 5" as shown below Fig. 7.

5.1.2 Establishment and solution of the "Fifth level" model

1. Determination of route

Considering only the shortest routes, the **3-day routes** are:

$$P_1 = \begin{pmatrix} 1,5,6,13 \\ 1,4,6,13 \end{pmatrix}$$

The **4-day routes** include:

$$P_2 = \begin{pmatrix} 1,4,7,12,13 \\ 1,4,6,12,13 \\ 1,4,7,11,13 \\ 1,5,6,12,13 \end{pmatrix}$$

Since this question involves mines, we also consider **5-day routes**, categorized as:

With mining opportunities (P_3):

$$P_3 = \begin{pmatrix} 1,2,3,9,10,13 \\ 1,2,3,9,11,13 \\ 1,4,3,9,10,13 \\ 1,4,3,9,11,13 \end{pmatrix}$$

Without mining opportunities (P_4):

$$P_4 = \begin{pmatrix} 1,4,7,11,12,13 \\ 1,5,6,7,12,13 \\ 1,5,6,12,11,13 \end{pmatrix}$$

Each row in P represents a distinct route.

2. Determination of the plan

Since the two players are in a competitive relationship, both players consider routes with less consumption or more benefits, so they only consider both choosing the three-day route, both choosing the four-day route, or one of them choosing the three-day route and the other choosing the four-day route, or both choosing the five-day mining route.

(1) Both players choose the 3-day route:

$$W = W_1 + W_2 = 2W_0 - 2\sum_{i=1}^{2} d_i(b_{iq} + b_{ig}) - 4\sum_{i=1}^{2} d_i b_{iq}$$

(2) Both players choose the 4-day route:

a. Choosing P_1 and P_2:

$$W = W_1 + W_2 = 2W_0 - 2\sum_{i=1}^{2} d_i(b_{iq} + b_{ig}) - 4\sum_{i=1}^{2} d_{iq}b_{ig}$$

b. Choosing P_1 and P_3:

$$W = W_1 + W_2 = 2W_0 - 2\sum_{i=1}^{2} d_{iq}b_{ig} - 4\sum_{i=1}^{2} d_i(b_{iq} + b_{ig})$$

c. Choosing P_1 and P_4:

$$W = W_1 + W_2 = 2W_0 - 2\sum_{i=1}^{2} d_i(2b_{iq} + b_{ig}) - 4\sum_{i=1}^{2} d_i b_{iq}$$

d. Choosing P_2 and P_4:

$$W = W_1 + W_2 = 2W_0 - 2\sum_{i=1}^{2} d_i(2b_{iq} + b_{ig}) - 4\sum_{i=1}^{2} d_i b_{ig}$$

e. Choosing P_3 and P_4:

$$W = W_1 + W_2 = 2W_0 - 2\sum_{i=1}^{2} d_i(2b_{iq} + b_{ig}) - 4\sum_{i=1}^{2} d_i b_{ig}$$

f. Choosing P_2 and P_4:

$$W = W_1 + W_2 = 2W_0 - 2\sum_{i=1}^{2} d_i(3b_{iq} + b_{ig})$$

(3) One chooses a 3-day route, the other chooses a 4-day route:

a. P_1 and P_2:

$$W = W_1 + W_2 = 2W_0 - 4\sum_{i=1}^{2} d_i(b_{iq} + b_{ig}) - 2\sum_{i=1}^{2} d_i b_{ig}$$

b. P_1 and P_2, P_3:

$$W = W_1 + W_2 = 2W_0 - 4\sum_{i=1}^{2} d_i b_{iq} - 2\sum_{i=1}^{2} d_i(b_{iq} + b_{ig})$$

c. P_1 and P_4:

$$W = W_1 + W_2 = 2W_0 - 4\sum_{i=1}^{2} d_i(b_{iq} + b_{ig}) - 2\sum_{i=1}^{2} 3d_i b_{iq}$$

d. Other situations:

$$W = W_1 + W_2 = 2W_0 - 2\sum_{i=1}^{2} d_i(b_{iq} + b_{ig})$$

(4) Both go mining, with the optimal strategy being P_2, P_3:

$$W = W_1 + W_2 = 2W_0 - 2\sum_{i=1}^{2} d_i(b_{iq} + b_{ig}) - 4\sum_{i=1}^{2} d_i b_{iq}$$

For "Scenario 3," the best option for both players is the above strategy, which yields the maximum value W.

3. Model solution

Use C++ to solve 12 situations and perform simulation tests. The results of the 12 cases are:

19080	18540	19020	19290	19180	19290
19400	18910	19180	18910	19290	19400

Use the RAND function and the system time as the random seed to generate random integers from 1 to 12. Use C++ to perform 100 simulations which preliminary analysis showed was sufficient for the mean outcome to stabilize and take the average as the simulation result, which is 10147.

5.2 Only know the weather of the day and the player's condition after the day ends

5.2.1 Strategies that players should generally adopt

Because after the end of each day's game, players will determine tomorrow's action plan based on their own funds and resources, and at the same time know other players' funds and resources to infer other players' action plans. In this case, players need to choose the movement plan that consumes the least funds and resources for the next day.

5.2.2 Specific Analysis and Discussion of the "Sixth Level"

Since the map of the sixth level is the same as that of the fourth level, and only the weather conditions of the day are known, players can determine the action plan for the next day based on the general strategy adopted by players in the face of unknown weather in question 2. Assume that the three players in the sixth level are in a competitive relationship with each other, and there is no alliance between players.

1. Introduce 0-1 variables to control the weather

$$x_1 + x_2 + x_3 = 1, \quad x_1, x_2, x_3 \in \{0, 1\}$$

2. Dynamic data prediction for players

Assume that the unknown location of the m person after the nth iteration is R_{mn}, and combine the map of the "sixth level" to predict the dynamic data of the player. Draw the route on the map of "Level 6" as shown below Fig. 8.

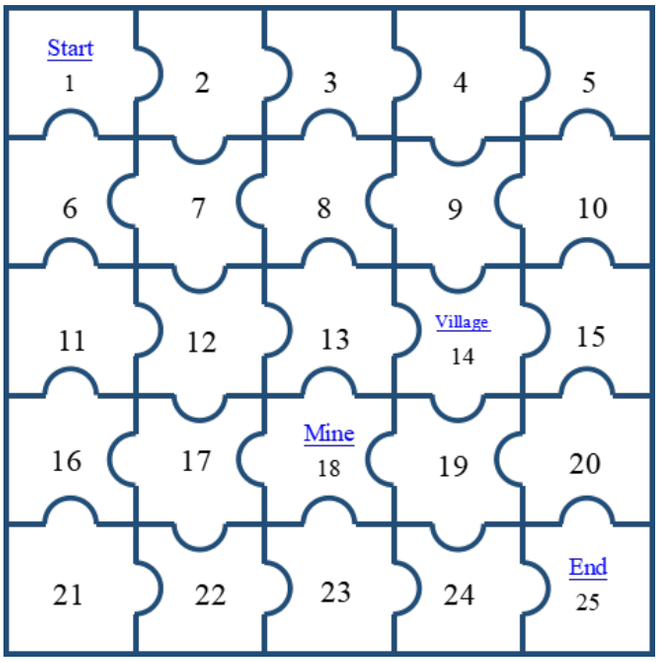

Figure 8: "Level 6" map

(1) Mining Constraints

a. After the n iteration, the player is around the mine:

$$|18 - R_{on}| = 5 \text{ or } 1, \quad |18 - R_{on}| = 5 \text{ or } 1, \quad a, b \in \{1, 2, 3\}$$

b. The player has enough funds to go to the mine and mine for at least one day:

$$W_{on} \geq 6 \sum_{i=1}^{2} (x_i b_{ip} d_i + x_i b_{ig} d_i + x_i b_{iq} d_i)$$

or

$$W_{on} \geq 6 \sum_{i=1}^{2} (x_i b_{ip} d_i + x_i b_{ig} d_i + x_i b_{iq} d_i), \quad a, b \in \{1, 2, 3\}$$

Players can go mining $R_{n(n+1)} = 18$ or $R_{n(n+1)} = 18$ if conditions a.b. are met.

(2) Path Restriction

When two players are in the same area and are ready to move to the next area:

$$|R_{on} - R_{on}| = 0$$

$$W_{on} \geq 2 \sum_{i=1}^{2} (x_i b_i p_i d_i + x_i b_i p_i d_i + x_i b_i p_i d_i)$$

$$W_{on} \geq 2 \sum_{i=1}^{2} (x_i b_i p_i d_i + x_i b_i p_i d_i + x_i b_i p_i d_i)$$

The two players do not move to the same area:

$$R_{n(n+1)} \neq R_{b(n+1)}$$

(3) Village Restrictions

a. After the n iteration, the player is around the village:

$$|14 - R_{on}| = 5 \text{ or } 1, \quad |14 - R_{on}| = 5 \text{ or } 1, \quad a, b \in \{1, 2, 3\}$$

b. The player has enough funds to go to the village, and must go to the village to replenish resources before moving:

$$2 \sum_{i=1}^{2} (x_i b_{ip} d_i + x_i b_{ig} d_i + x_i b_{iq} d_i) \leq W_{an}$$

$$W_{an} \leq 4 \sum_{i=1}^{2} (x_i b_{ip} d_i + x_i b_{ig} d_i + x_i b_{iq} d_i)$$

$$\text{or} \quad 2 \sum_{i=1}^{2} (x_i b_{ip} d_i + x_i b_{ig} d_i + x_i b_{iq} d_i) \leq W_{bn}$$

$$W_{bn} \leq 4 \sum_{i=1}^{2} (x_i b_{ip} d_i + x_i b_{ig} d_i + x_i b_{iq} d_i), \quad a, b \in \{1, 2, 3\}$$

(4) Competition between players who reach a special point at the same time

a. Arrive at the mine at the same time. Assume that people with sufficient funds will enter the mine to mine at this time, that is , players with less funds are unwilling to halve the mining benefits (players will lose money).

b. When they arrive at the village at the same time, those who do not have enough resources will definitely enter the village, while those who have enough resources will not enter the village.

c. Players all consider their own interests in the game. If the level differences among players are not big, the final value will be smaller after the game ends.

6 Conclusions

1. Advantages

(1) Using dynamic programming to solve the knapsack problem. Dynamic programming has memory, and the subproblems needed in the new problem can be directly extracted, avoiding repeated calculations and thus saving time.

(2) The knowledge of game theory is used to simplify the abstract concepts of the fifth and sixth levels to facilitate the solution and analysis of random problems.

(3) Converting graph theory problems into dynamic optimization problems provides a structured and computable framework for solving complex routing decisions. This integration greatly enhances the practicality of the model by transforming an intuitive but vague map-based decision into a series of well-defined optimization steps with clear objectives and constraints. The process is illustrated in Figure 9, which shows how the spatial problem (the map and weather) is abstracted into a graph, then into a state-space model, and finally solved via DP to yield an optimal policy. This methodological pipeline is a core contribution of our work.

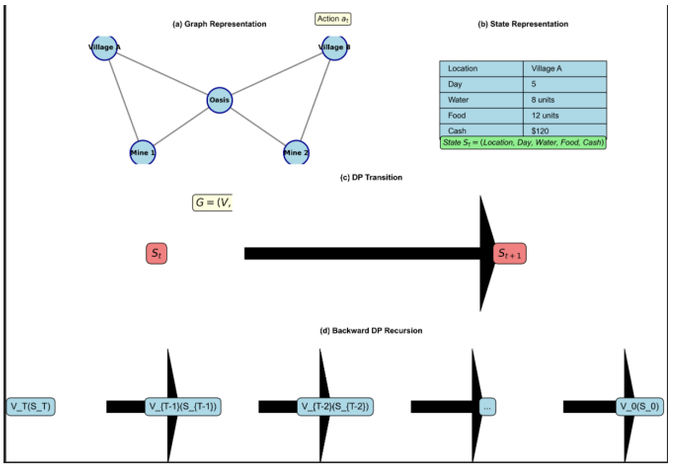

Figure 9: Modeling Framework: From Graph Theory to Dynamic Optimization

2. Disadvantages:

(1) There is a lack of empirical validation of model results against alternative strategies (e.g., greedy algorithms) or real-world data, which results in uncertain error margins compared with actual processes or simpler benchmarks. Future work should include such validation.

(2) In the cycle analysis of mines and villages, the accuracy of the solution is reduced due to the unknown number of cycles. Heuristics were used to approximate the optimal number.

(3) The models are developed specifically for the desert game's ruleset (specific resources, weather, map structures). Generalizing the framework to other domains (e.g., logistics with fuel constraints, multi-robot exploration with charging stations) would require adapting the cost functions, constraints, and state definitions, but the core methodology of combining DP for planning, graph search for routing, and game theory for interaction remains promising. Future research should explore this generalization.

(4) Scalability to larger maps or more players needs assessment, potentially requiring more efficient algorithms or approximations.

References

[1] Contemporary undergraduate mathematical contest in modeling (cumcm). Accessed 26 April 2021.

[2] Altman, E., and Wynter, L. Equilibrium, games, and pricing in transportation and telecommunication networks. *Networks and Spatial Economics* **4** (2004), 7–21.

[3] Bellman, R. *Dynamic Programming*. Princeton University Press, Princeton, NJ, 1957.

[4] Cao, Z. Exploration of exhaustive, search, and dynamic programming algorithms for the 0-1 knapsack problem. *Computer Knowledge and Technology* **12** (2009), 3193–3194.

[5] Carpin, S. Solving stochastic orienteering problems with chance constraints using monte carlo tree search. *arXiv:2409.03170*, 2024.

[6] Dijkstra, E. W. A note on two problems in connexion with graphs. *Numerische Mathematik* **1**(1) (1959), 269–271.

[7] Tang, J., Liu, Y., and Zhu, X. Reinforcement learning approaches for the orienteering problem. *Transportation Science* (2022). Forthcoming.

[8] Xu, J., Liu, K., and Liu, J. Competitive knapsack routing: A nash-dp approach. *Computers & Operations Research* **135** (2021), 113–125.

[9] Zhang, F., Liu, J., and Li, Q. An optimized shortest-path algorithm based on dijkstra's method. *Remote Sensing Information* **000**(002) (2004), 38–41.

[10] Zhang, Y., Wang, L., and Li, D. Data-driven time-dependent orienteering with stochastic travel times. *Computers & Operations Research* **145** (2023), 105–123.

7 List of Appendix

Appendix 1: List of all supporting documents
Appendix 2: "Level 1" description matrix
Appendix 3: LINGO program for solving the "first level"
Appendix 4: Results table of the "first level"
Appendix 5: "Level 2" description matrix
Appendix 6: Results table of the "Second Level"
Appendix 7: "Level 3" description matrix
Appendix 8: Descriptive matrix for the "Fourth Level"
Appendix 9: LINGO program for solving the fourth level
Appendix 10: LINGO program for solving the fifth level

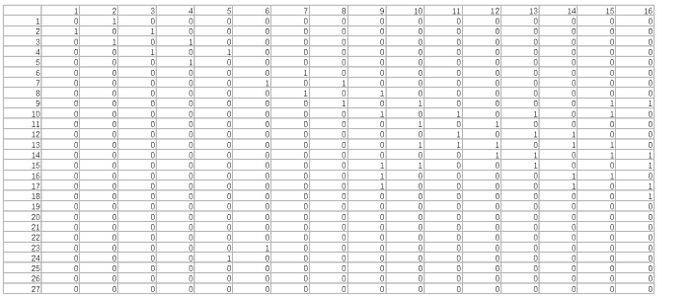

Figure 10: "Level 1" description matrix

Sets:

$$a = \{1, 2\}$$
$$a_1 = \{1, 2, 3\}$$
$$b = \{1, 2, 3, 4\}$$
$$\text{link}(a, b)$$

Data:

$$d = [5, 10], \quad bq = [8, 6], \quad bg = [5, 7], \quad bs = [10, 10],$$
$$w = [5, 10], \quad p = 1000$$

Initialization:

$$fv(1) = 0$$

Revenue and Resource Use Definitions:

$$rw = rw_1 + rw_2$$
$$rw_1 = \sum_{i \in a} d_i \cdot k_i$$
$$rw_2 = 2 \sum_{i \in a} d_i \cdot k_{i+2}$$
$$rw_1 < w_0$$
$$rw_2 < w_0 - rw_1 + p \cdot (t_4 - 1)$$

Capacity Constraint:

$$\sum_{i \in a} c_i \cdot (k_i + k_{i+2}) - \sum_{i \in a} (x_{i1} + x_{i4}) \leq M_1$$

Total Time at Each Stage:

$$t_j = tg_j + tq_j + ts_j, \quad \forall j \in b$$

Resource Consumption:

$$x_{ij} = \begin{cases} 3 \cdot (bg_i \cdot tg_j + bq_i \cdot tq_j + bs_i \cdot ts_j), & \text{if } j = 4 \\ 2 \cdot (bg_i \cdot tg_j + bq_i \cdot tq_j) + bs_i \cdot ts_j, & \text{otherwise} \end{cases}$$

Residual Value:

$$rwp = \frac{1}{2} \sum_{i \in a} d_i \cdot \left(k_i + k_{i+2} - \sum_{j \in b} x_{ij} \right)$$

Forward Value Computation:

$$fv_{i+1} = \max\left(fv1_i + k_i \cdot w_i\right), \quad \forall i \in a$$

Variable Constraints:

$$k_i \leq m_i, \quad \forall i \in a$$
$$k_i - \sum_{j \in \{1,2,3\} \cup \{4\}} x_{ij} > 0, \quad \forall i \in a$$

Objective Function (to Maximize):

$$\max Z = w_0 + p \cdot (t_4 - 1) + rwp - rw$$

0	1	5800	180	330
1	25	5800	164	318
2	24	5800	148	306
3	23	5800	138	292
4	23	5800	128	282
5	22	5800	118	268
6	9	5800	102	256
7	9	5800	92	246
8	15	4170	245	232
9	14	4170	229	220
10	12	4170	213	208
11	12	5170	183	178
12	12	6170	159	160
13	12	7170	144	139
14	12	8170	120	121
15	12	9170	96	103
16	12	10170	72	85
17	12	11170	42	55
18	12	11170	32	45
19	14	11170	16	33
20	15	10430	36	40
21	9	10430	26	26
22	21	10430	16	12
23	27	10430	0	0
24				
25				
26				
27				
28				
29				
30				

Figure 11: Results table of the “first level”

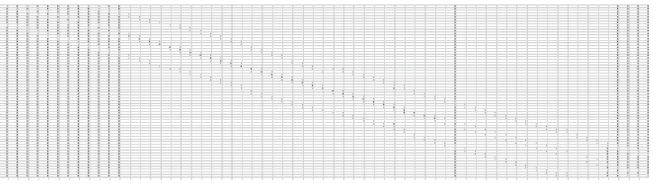

Figure 12: ”Level 2” description matrix

Sets:

$a = \{1, 2\}$: parameters d, bq, bg, bs, m
$b = \{1, 2, 3\}$
$b_1 = \{1, 2, 3, 4\}$: with $k(i)$
$c = \{1, 2, \ldots, 9\}$: with time variables $t(n)$
link$(a, b_1) : x(i, j)$

Data:

$w_0 = 10000$
$d = [5, 10]$
$bq = [3, 4]$
$bg = [9, 9]$
$bs = [10, 10]$
$p = 1000$

0	1	6475	247	229
1	2	6475	231	217
2	3	6475	215	205
3	4	6475	205	191
4	4	6475	195	181
5	5	6475	185	167
6	13	6475	169	155
7	13	6475	159	145
8	22	6475	149	131
9	30	6475	133	119
10	30	7475	109	101
11	30	8475	79	71
12	30	9475	55	53
13	30	10475	40	32
14	39	11475	16	14
15	46	6775	241	237
16	55	4365	225	225
17	55	4365	215	215
18	55	4365	205	205
19	55	4365	189	193
20	55	7775	165	175
21	55	8775	150	154
22	55	9775	135	133
23	55	10775	111	115
24	55	11775	96	94
25	55	9365	86	84
26	55	12775	62	66
27	55	13775	47	45
28	55	14775	32	24
29	63	12365	16	12
30	64	12365	0	0

Figure 13: Results table of the “Second Level”

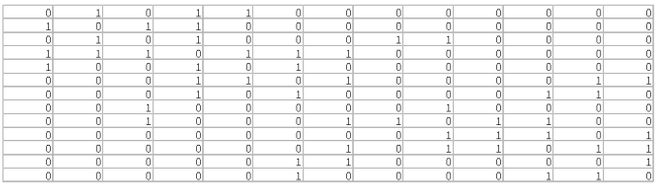

Figure 14: ”Level 3” description matrix

(1) Total supply constraint:

$$\sum_{i \in a} m(i) < 1200$$

(2) Time assignment for last 3 periods:

$$\sum_{\substack{n \in c \\ n \geq 7}} t(n) = 8$$

(3) First-stage remaining wealth:

$$w_1 = w_0 - \sum_{i \in a} d(i) \cdot k(i)$$

(4) Total working time constraint:

$$\sum_{n \in b} t(n) \leq 30$$

(5) Maximum mining quantity:

$$k(i) \leq m(i), \quad \forall i \in a$$

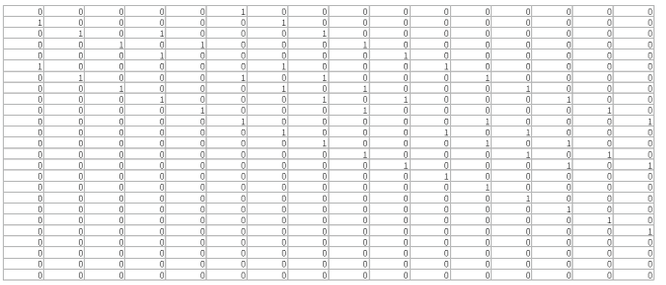

Figure 15: Descriptive matrix for the ”Fourth Level”

(6) Resource availability constraint (stage-wise):

$$k(i) - 2\left[bq(i)(t(1) - t(4)) + bg(i)(t(2) - t(5))\right] - bs(i)(t(3) - t(6)) - 3\left[bq(i)t(4) + bg(i)t(5) + bs(i)t(6)\right] \geq 0, \quad \forall i \in a$$

(7) Wealth after mining and profit:

$$w_1 = w_0 - \sum_{i \in a} d(i) \cdot k(i) + p \cdot (t(4) + t(5) + t(6) - 1)$$

(8) First player’s individual constraint:

$$k(1) + k(3) - 2 \cdot (bg(1) + bq(1) - bs(1) \cdot t(3)) \geq 0$$

(9) Second player’s individual constraint:

$$k(2) + k(4) - 2 \cdot (bg(2) + bq(2) - bs(2) \cdot t(3)) \geq 0$$

(10) Resource consumption calculation:

$$x(i,j) = 2 \cdot (bq(i)t(j) + bg(i)t(j) + bs(i)t(j)), \quad \forall (i,j) \in \text{link}(a, b_1)$$

$$\max Z = w_0 - \sum_{i \in a} d(i) \cdot k(i) - \sum_{i \in a} d(i) \cdot k(i+2)$$

$$+ \frac{1}{2} \sum_{i \in a} \left[k(i) + k(i+2) - \sum_{j \in b_1} x(i,j) \right]$$

Listing 1: C++ Program for Mining Strategy Evaluation

```
#include<iostream>
#include<iomanip>
#include<cstdlib>
#include<ctime>
#define random(a,b) (rand()%(b-a)+a)

using namespace std;

int sigma1(int *a, int *b, int *c)
{
    int result = 0;
    for (int i = 0; i < 2; i++)
    {
        result = result + (b[i] + c[i]) * a[i];
    }
```

```
    return result;
}

int sigma2(int *a, int *b)
{
    int result = 0;
    for (int i = 0; i < 2; i++)
    {
        result = result + a[i] * b[i];
    }
    return result;
}

int main()
{
    int f[100], sum = 0, av;
    ((int)time(0)); // Generate random seeds. Replace 0 with NULL and you will see 2 rows.

    for (int i = 0; i < 100; i++)
    {
        f[i] = (rand() % (11 - 0 + 1)) + 0;
    }

    int bg[2] = {9, 9}, bq[2] = {3, 4}, bq2[2] = {6, 8}, bq3[2] = {9, 12}, bq4[2] = {12, 16}
    int w0 = 10000, w[20], d[2] = {5, 10};

    w[0]  = 2 * w0 - sigma1(d, bg, bq) * 2 - 4 * sigma2(bg, d);
    w[1]  = 2 * w0 - sigma1(d, bg, bq) * 2 - 8 * sigma2(bg, d);
    w[2]  = 2 * w0 - sigma1(d, bg, bq) * 4 - 4 * sigma2(bq, d);
    w[3]  = 2 * w0 - sigma1(d, bg, bq2) * 2 - 4 * sigma2(bq, d);
    w[4]  = 2 * w0 - sigma1(d, bg, bq) * 2 - 4 * sigma2(bq2, d);
    w[5]  = 2 * w0 - sigma1(d, bg, bq2) * 2 - 4 * sigma2(bq, d);
    w[6]  = 2 * w0 - sigma1(d, bg, bq3) * 2;
    w[7]  = 2 * w0 - sigma1(d, bg, bq) * 4 - 2 * sigma2(bq3, d);
    w[8]  = 2 * w0 - sigma1(d, bg, bq3) * 2 - 4 * sigma2(bq, d);
    w[9]  = 2 * w0 - sigma1(d, bg, bq) * 4 - 2 * sigma2(bq3, d);
    w[10] = 2 * w0 - sigma1(d, bg, bq4) * 2;
    w[11] = 2 * w0 - 2 * sigma1(d, bg, bq) - 4 * sigma2(d, bq);

    for (int i = 0; i <= 11; i++)
        cout << w[i] << '\t';

    for (int k = 0; k < 100; k++)
    {
        sum = sum + w[f[k]];
        av = sum / (k + 1);
    }

    cout << av;
    return 0;
}
```

In Silico Cloning and Bioinformatics Analysis of Shikimate Dehydrogenase Gene from *Medicago sativa*

GuanPing You[1, 2] and JianZhong Huang[1, 2]
[1] College of Life Sciences, Fujian Normal University, Fuzhou, Fujian, China
[2] Engineering Research Center of Industrial Microbiology, Ministry of Education, National and Local United Engineering Research Center of Industrial Microbiology and Fermentation Technology, Fujian Normal University, Fuzhou, Fujian, China

Abstract—The combination of artificial intelligence (AI) and bioinformatics is driving a leap forward in genomics and biological research, especially in the electronic cloning and biological analysis of genes. AI can analyze large-scale genomic data, identify gene variations and predict gene functions through machine learning algorithms, thereby improving the efficiency and accuracy of gene cloning. Electronic cloning technology combines computer modeling and experimental data to simulate the gene expression process, greatly accelerating the progress of gene function research. In the secondary metabolic pathway of plants, shikimate dehydrogenase (SDH) is one of the key enzymes involved in the regulation of the shikimate pathway, which is a key step in the synthesis of important plant secondary metabolites such as phenylpropene compounds, flavonoids and lignin. Shikimate dehydrogenase catalyzes the conversion of shikimate to coumaric acid, which is the basis of plant defense mechanisms, antioxidants and disease resistance. In this study, AI tools were used to deeply analyze the gene expression patterns related to shikimate dehydrogenase, and the shikimate dehydrogenase sequence gene of *Escherichia coli* was used as a probe to clone and analyze the *Medicago sativa* shikimate dehydrogenase gene. The results showed that the cloned shikimate dehydrogenase gene of *M. sativa* was 469 bp in length and had 5 open reading frames (ORFs), of which ORF3 was the longest, with a total length of 258 bp, encoding 85 amino acids. The molecular weight of the protein was 9370.70, and the theoretical isoelectric point pI was 5.67, indicating that it was a functional protein on abiotic membranes. Through further bioinformatics analysis, it was speculated that the gene may play an important role in the secondary metabolism of *M. sativa*, and its expression pattern may be closely related to the growth and environmental adaptability of the plant.

Index Terms—Artificial intelligence, bioinformatics, *Medicago sativa*, shikimate dehydrogenase, in silico cloning

I. INTRODUCTION

Shikimate dehydrogenase (SDH) is a key enzyme in the shikimate pathway, which plays a vital role in the synthesis of aromatic amino acids and their precursors in plants. This pathway is central to plant secondary metabolism, contributing not only to the production of amino acids such as tryptophan, tyrosine, and phenylalanine, but also to certain plant hormones and a vast array of secondary metabolites (Fig. 1). Beyond metabolism, the shikimate pathway influences plant growth, environmental adaptation, stress resistance, and yield. Its industrial significance is also notable, particularly as shikimic acid serves as a key precursor for the synthesis of the antiviral drug oseltamivir (Tamiflu)[1]. Genetic engineering of this pathway offers promising avenues for enhancing the production of valuable compounds and improving agronomic traits[1].

Medicago sativa (alfalfa), an important forage and green manure crop, exhibits shikimate pathway activity that is closely linked to stress tolerance and biomass yield. Although recent genetic studies on *M. sativa* have increasingly focused on stress-related genes, functional characterization of SDH — a pivotal enzyme in secondary metabolism — has remained limited[2]. The integration of artificial intelligence (AI) with bioinformatics has created new opportunities for accelerating gene discovery and functional prediction[3-4]. Yet, current approaches often rely on generalized bioinformatics tools that lack custom analysis tailored to non-model species such as alfalfa, and most conventional methods do not fully leverage AI for predictive functional insight.

To address these limitations, this study employs a targeted in silico cloning strategy combined with multi-level bioinformatics analysis to identify and characterize the SDH gene from *M. sativa*. We present the first report of a putative SDH gene in alfalfa, comprising 469 bp with five open reading frames (ORFs), among which ORF3 encodes an 85-amino acid protein. Our analysis reveals key protein characteristics including a molecular weight of 9370.70, an acidic isoelectric point (pI) of 5.67, hydrophilic nature, and cytoplasmic localization supported by the absence of signal peptides and transmembrane domains. Furthermore, functional motif identification and phylogenetic analysis provide insight into the evolutionary conservation and functional role of SDH in legumes.

These findings establish a essential genetic resource for future research on metabolic engineering in alfalfa, and demonstrate a bioinformatics workflow that can be augmented with AI tools for improved gene function prediction. This work not only facilitates further functional validation of SDH but also

This work was supported by the National Key Research and Development Program of China under Grant No. 2022YFD1802104. Corresponding author: JianZhong Huang (e-mail: hjz@fjnu.edu.cn). The author GuanPing You(e-mail: 1257887082@qq.com) is with the College of Life Sciences, Fujian Normal University, Fuzhou, Fujian, China.

provides a foundation for enhancing stress resistance and yield in *M. sativa* through molecular breeding.

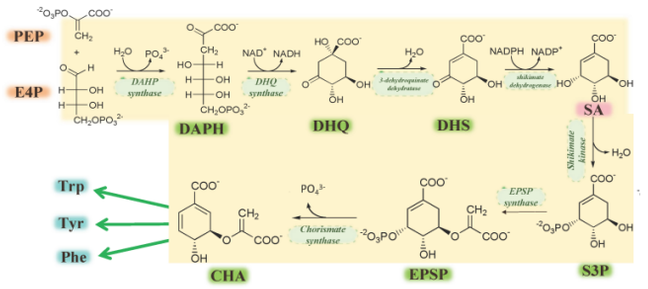

Fig. 1. The shikimate pathway. PEP: Phosphoenolpyruvate; E4P: D-erythrose 4-phosphate; DAHP: 3-deoxy-d-arabino-heptulosonic acid 7-phosphate; DHQ: 3-dehydroquinate; DHS: 3-dehydroshikimate; SA: Shikimate; S3P: Shikimate-3phosphate; EPSP: 5-enolpyruvylshikimate-3-phosphate; CHA: Chorismate; Phe: Phenylalanine; Tyr: Tyrosine; Trp: Tryptophan.

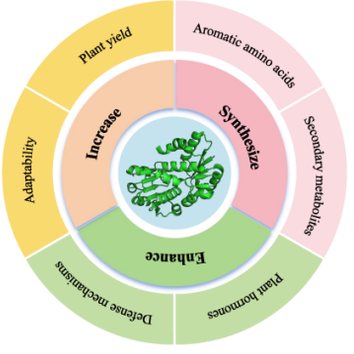

Fig. 2. Application of shikimic acid in plants.

II. Related Work

The shikimate pathway is a fundamental metabolic route in plants, bacteria, and fungi, serving as the crucial bridge between primary carbon metabolism and the biosynthesis of aromatic amino acids and a vast array of secondary metabolites. Within this pathway, shikimate dehydrogenase (SDH, EC 1.1.1.25) catalyzes the reversible reduction of 3-dehydroshikimate to shikimate, utilizing NADPH as a cofactor. This reaction is a critical control point, making SDH a subject of significant interest in both basic research and applied biotechnology.

A.Studies on SDH Genes in Various Species

Extensive research has been conducted on SDH genes across different kingdoms. In bacteria, such as *Escherichia coli,* the *aroE* gene encoding SDH has been well-characterized, and its structure-function relationship has been elucidated, providing a foundational model for understanding enzyme kinetics and mechanism. In plants, SDH has been identified and studied in model species like *Arabidopsis thaliana* and major crops such as rice (Oryza sativa). These studies have confirmed SDH's pivotal role in development and stress responses. For instance, the silencing of SDH in *Arabidopsis* led to severe developmental defects, underscoring its indispensability. Furthermore, recent work has begun to explore SDH in bioenergy crops like Panicum virgatum (switchgrass), highlighting its potential in engineering pathways for improved biomass and stress resilience. While these studies provide a general framework for understanding SDH, functional characteristics can vary significantly between species due to evolutionary divergence and lineage-specific adaptations.

B. Research Gap and Objective

While SDH is recognized as a critical enzyme, its specific sequence, structure, and functional attributes in *Medicago sativa*, a legume crop of immense agricultural importance, remain poorly characterized. Previous studies in other species provide a template but cannot directly be extrapolated. The conventional bioinformatics approaches used in many prior SDH studies, while useful, lack the predictive power of modern AI techniques. Therefore, there is a clear need for a comprehensive study that not only identifies and characterizes the *M. sativa* SDH gene using established *in silico* methods but also frames these findings within the modern context of AI-driven biological discovery. This study aims to fill this gap by conducting a foundational bioinformatic characterization of *M. sativa* SDH and explicitly outlining a pathway for its future validation and application using advanced computational intelligence, thereby contributing to the genetic improvement of this vital crop.

III. Analysis Methods and Tools

A. In silico cloning of SDH from M. sativa

The probe amino acid sequence of the *E. coli* SDH used in this study is following:

METYAVFGNPIAHSKSPFIHQQFAQQLNIEHPYGRVL APINDFINTLNAFFSAGGKGANVTVPFKEEAFARADELT ERAALAGAVNTLMRLEDGRLLGDNTDGVGLLSDLERL SFIRPGLRILLIGAGGASRGVLLPLLSLDCAVTITNRTVS RAEELAKLFAHTGSIQALSMDELEGHEFDLIINATSSGIS GDIPAIPSSLIHPGIYCYDMFYQKGKTPFLAWCEQRGSK RNADGLGMLVAQAAHAFLLWHGVLPDVEPVIKQLQEE LSA.

Next, the probe sequence was searched with the M. sativa EST database using the tBlastn tool in NCBI, and gene sequences with a match greater than 50% were selected and downloaded in FASTA format; contig-0 and contig-1 were obtained by gene splicing using BioEdit "CAP conting assembly program". Finally, the *M. sativa* shikimate dehydrogenase cDNA sequence was predicted by ORF Finder in NCBI to determine whether there was a gene with the expected function, and finally the new gene fragment was determined. The analysis tools are shown in Table 1.

TABLE I
TOOLS USED TO PREDICT GENE STRUCTURE

Search content	Tools
Sequence acquisition	NCBI

	https://www.ncbi.nlm.nih.gov/
Sequence splicing	BioEdit
Open reading frame identification	ORFfinder https://www.ncbi.nlm.nih.gov/orffinder/

B. Bioinformatics Analysis of the SDH in M. sativa

The obtained *M. sativa* SDH gene sequence was analyzed using bioinformatics software(*Sequence alignments were auto-assembled via BioEdit CAP, with manual curation of ambiguous regions.* , and an evolutionary tree was constructed by analyzing and predicting the physicochemical properties, hydrophilicity and hydrophobicity, functional sites, transmembrane analysis, signal peptides, subcellular localization, secondary structure, tertiary structure, molecular evolution, etc. of the protein encoded by the gene. The specific analysis content and tool software are shown in Table 2.

TABLE II
TOOLS USED TO PREDICT PROTEIN STRUCTURE AND FUNCTION

Search content	Tools and Parameters
Base composition	Bioedit（Nucleotide Composition, Restriction Map）
Physical and chemical properties	http://web.expasy.org/protparam/
Hydrophilicity/hydrophobicity	http://web.expasy.org/protscale/
Functional sites	https:/web.expasy.org/ prosite/
Transmembrane analysis	TMHMM2.0,Default settings (membrane probability >0.5; N-tail inside) https://services.healthtech.dtu.dk/services/TMHMM-2.0/
Subcellular localization	https://wolfpsort.hgc.jp/
Signal peptide	https://services.healthtech.dtu.dk/service.php?SignalP-5.0
Secondary structure	SOPMAD,Window width=17; Decision constants: Helix (≥4), Sheet (≥4) https://npsa-prabi.ibcp.ft/cgi-bin/npsaautomat.pl?page=npsa%20sopma.html
Phosphorylation site	https://services.healthtech.dtu.dk/services/NetPhos-3.1/
Tertiary structure	http://swissmodel.expasy.org/repository/
Homologous evolutionary tree	MEGA11,Neighbor-Joining (NJ) tree; Bootstrap=1000 replicates; Poisson correction

III. AMINO ACID COMPOSITION ANALYSIS OF SDH IN *M. SATIVA*

Amino acid	Number	Proportion	Amino acid	Number	Proportion
Ala(A)	9	10.6%	Ile(I)	7	8.2%
Arg(R)	3	3.5%	Leu(L)	8	9.4%
Asn(N)	2	2.4%	Lys(K)	3	3.5%
Asp(D)	5	5.9%	Met(M)	3	3.5%
Cys(C)	3	3.5%	Phe(F)	4	4.7%
Gln(Q)	3	3.5%	Pro(P)	4	4.7%
Glu(E)	4	4.7%	Ser(S)	7	8.2%
Gly(G)	8	9.4%	Thr(T)	2	2.4%
His(H)	4	4.7%	Trp(M)	2	2.4%
Val(V)	1	1.2%	Tyr(Y)	3	3.5%

A. In silico cloning results of shikimate dehydrogenase from M. sativa

The EST sequences obtained by TBLASTN were saved in FASTA format. The sequences were spliced using Bioedit software to finally obtain a contig with a length of 469bp. Then, through the prediction tool ORF Finder in NCBI, the *M. sativa* SDH sequence has 5 ORFs, of which the longest ORF is 258bp long. The protein it encodes contains 86 amino acids, and the sequence is:

MDELEGHEFDLIINATSSGISGDIPAIPSSLIHPGIYCYD MFYQKGKTPFLAWCEQRGSKRNADGLGMLVAQAAHA FLLWHRCSA. Sequences with high homology were found through BLAST. Using the Bioedit tool, the base composition can be obtained by analyzing the electronic cloning splicing sequence, among which the proportion of adenine A is 28.14%; the proportion of cytosine C is 27.08%; the proportion of guanine G is 22.17%; and the proportion of thymine T is 22.60%. Analysis of restriction enzyme positions of the electronically cloned spliced sequence revealed that the restriction enzyme sites of the nucleotide sequence of the *M. sativa* SDH include AflIII, AlwI, AseI, and the like.

B. Analysis of the physicochemical properties of M. sativa shikimate dehydrogenase protein

The ExPASy-Protparam tool was used to analyze the physicochemical properties of *M. sativa* shikimate dehydrogenase. The analysis showed that the number of amino acids was 85, the molecular weight of the protein was 9370.70, the theoretical isoelectric point pl was 5.67, the molecular formula was $C_{419}H_{636}N_{112}O_{121}S_6$, the instability coefficient was 35.85, it was a stable protein, the fat coefficient was 82.82, and the average hydrophobicity was -0.031. Its amino acid composition is shown in Table 3. There are 9 negatively charged amino acids (Asp + Glu), 6 positively charged amino acids (Arg + Lys), Ala accounts for the largest proportion, 10.6%, there is no pyrrolysine (Pyl) and selenocysteine (Sec), and the protein is an acidic protein.

C. Prediction and analysis of hydrophilicity/hydrophobicity of M. sativa shikimate dehydrogenase

The ExPASy-ProtScale online software was used to predict

the hydrophilicity/hydrophobicity of the amino acid protein encoded by the *M. sativa* SDH gene. The results are shown in Figure 3. Analysis of the figure shows that the larger the negative value, the weaker the hydrophilicity of the protein; conversely, the larger the positive value, the stronger the hydrophobicity of the protein. The value between +2 and -3 indicates that the amino acid is amphoteric. The highest score of the polypeptide chain is at the 69th position of the protein, which is +1.689, and the lowest score is at the 58th and 59th positions of the protein, which is -2.533, so it is a hydrophilic protein. It can also be observed that the negative peak is significantly higher than the positive peak, and it is inferred that the *M. sativa* shikimate dehydrogenase is hydrophilic. This is consistent with the analysis results of the ExPASy-Protparam software.

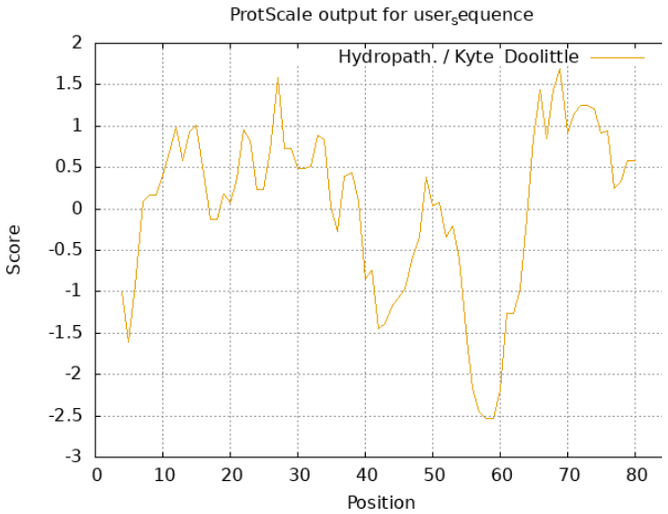

Fig. 3. The hydrophilicity/hydrophobicity prediction results of *M. sativa* SDH.

D. Prediction and analysis of signal peptide and transmembrane domain of M. sativa SDH

The signal peptide is a peptide segment consisting of 20 to 30 amino acid residues at the N-terminus of the nascent peptide chain of a secretory protein. It determines the modification of certain amino acid residues and is often used to guide the transmembrane transfer of proteins. The signal peptide structure of *M. sativa* SDH was predicted using the Signalp5.0 online software. The prediction results are shown in Figure 4. The results show that the probability of the protein being a signal peptide is 0.0015, and it can be inferred that the protein is a non-secretory protein，and the absence of signal peptides (SignalP5.0 score=0.0015) and transmembrane domains (TMHMM) confirms cytosolic localization. This aligns with SDH's role in cytoplasmic shikimate metabolism [1,3] and suggests direct interaction with cytoplasmic substrates like shikimate.

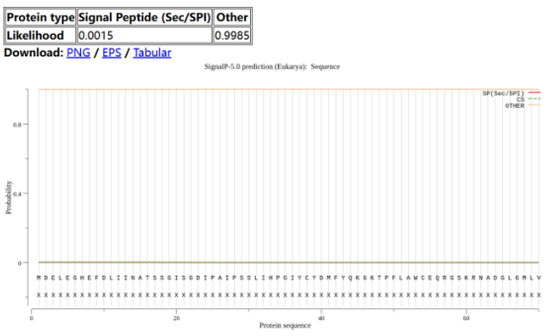

Fig. 4. Protein signal peptide prediction results of *M. sativa* SDH.

The transmembrane domain is the main part where the membrane-intrinsic protein binds to the membrane lipids. It is generally composed of about 20 hydrophobic amino acids to form an alpha helix, which is fixed to the cell membrane and acts as an anchor. The TMHMM-2.0 online tool was used to predict the transmembrane domain of the protein. The results are shown in Figure 5. At 1.0, it is the outer boundary of the cell membrane, and 0 is inside the cell membrane. This study predicted a protein sequence with a length of 85 amino acids, but no predicted transmembrane helices (TMHs) were found, indicating that the protein is likely to contain no transmembrane regions, or these regions may be too short or do not meet the prediction criteria of the TMHMM model. Overall, this protein may be a non-transmembrane protein or located inside the cell. Further, the protein was annotated and the functional site prediction was performed using ExPASy-Prosite. The results showed that the 14-17, 58-63, and 67-72 amino acids were the predicted functional sites of the protein.

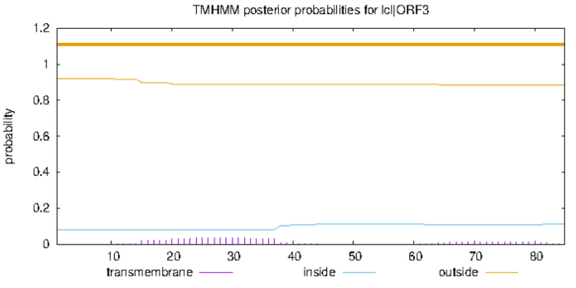

Fig. 5. Protein transmembrane domain prediction results of *M. sativa* SDH.

E. Structural prediction and analysis of M. sativa SDH

The local spatial structure of the polypeptide main chain of *M. sativa* SDH was analyzed using the online software SOPMA. The results showed that in the secondary structure of the protein, the largest proportion was random coils, accounting for 48.24%, with 41; followed by α helices, accounting for 31.76%, with 27; there were 17 extended chains, accounting for 20%; there was no β fold.

The SWISS-MODEL online homology modeling method was used to analyze the *M. sativa* SDH protein, predict its tertiary structure, and use the ball-and-stick model to display all aromatic amino acids. The results are shown in Figure 6, A is the predicted image of the tertiary structure of the protein, and the blue part shown in B is the position of the aromatic amino acids in the protein stick-and-ball model.

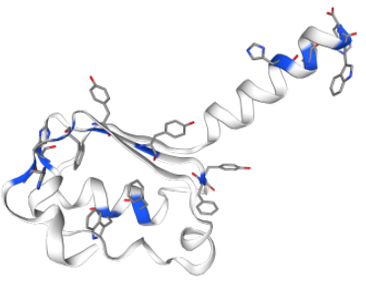

Fig. 6. Protein tertiary structure and predicted position of aromatic amino acids.

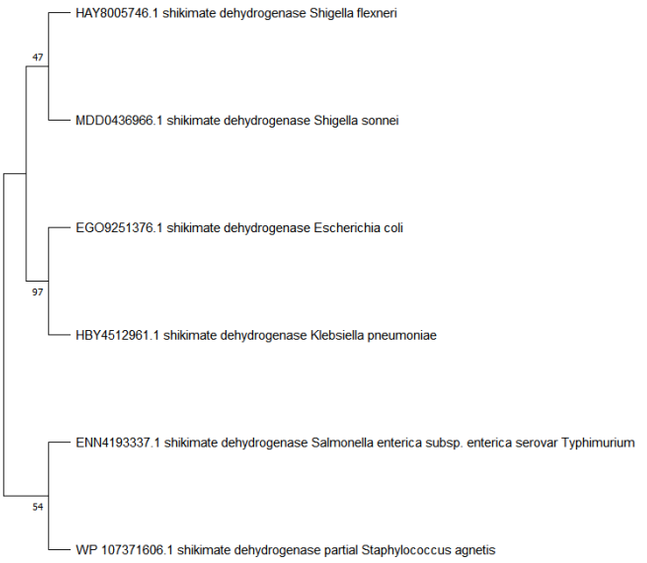

Fig. 7. Phylogenetic analysis of *M. sativa* SDH.

F. Construction of the phylogenetic tree of M. sativa SDH

Using the results from ORF Finder in NCBI, we obtained homologous sequences of various species through BLAST analysis, and then used the software MEGA11 to analyze the evolutionary tree structure (Figure 7). We searched for protein sequences similar to the protein sequence and performed multiple sequence alignment on them to construct the molecular evolutionary phylogenetic tree of the protein in plant species (using the neighbor-joining method, NJ).

Phylogenetic analysis indicates a clear evolutionary relationship among the SDH genes from the bacterial species examined. Sequences from members of the Enterobacteriaceae family—including *Escherichia coli* , *Shigella flexneri* , *Shigella sonnei* , *Klebsiella pneumonia*e , and *Salmonella enterica* — form a tightly clustered monophyletic clade with strong bootstrap support (values of 97 and 47), demonstrating high sequence similarity and suggesting a relatively recent common ancestor. In contrast, *Staphylococcus agnetis* (phylum Firmicutes) represents a more distantly related lineage. Its position as an outer group, with a bootstrap value of 54 at the divergent node, is consistent with its taxonomic distinction from the Enterobacteriaceae, underscoring the divergence between these bacterial groups.

IV. CONCLUSION

This study focused on the SDH of *M. sativa* and successfully obtained the full-length sequence of the gene using electronic cloning technology. Through a series of bioinformatics analysis tools and methods, the biological characteristics of the gene and the protein it encodes were systematically analyzed[5]. The results of gene sequence analysis revealed the characteristics of its base composition, potential restriction sites and sequence variation information, laying the foundation for subsequent gene function research and application. In terms of protein sequence analysis, we deeply explored key factors including functional sites, domains, physicochemical properties, hydrophilic and hydrophobic properties. In addition, signal peptide prediction, transmembrane structure analysis ~~and subcellular localization analysis~~ also provide an intuitive understanding of the distribution and action location of the protein in the cell. Further secondary and tertiary structure predictions provide a spatial structural basis for revealing the functional mechanism of the protein, while the phylogenetic tree constructed based on multiple sequence alignment and molecular evolution analysis reveals the evolutionary relationship of the gene, providing important clues for inferring its evolutionary process and genetic differences between species.

These research results not only provide key data support for our in-depth understanding of the SDH mechanism of *M. sativa* growth and development, but also provide a theoretical basis for the gene function and metabolic regulation of *M. sativa*. By revealing the specific function of this gene in the metabolic pathway of *M. sativa*, these data provide an important reference for *M. sativa* variety improvement, metabolic regulation and stress resistance research. At the same time, the gene sequence and protein characteristic analysis obtained in this study

provide a solid theoretical basis for subsequent functional verification experiments, and provide new ideas and directions for further exploring the application potential of *M. sativa* in agricultural production.

V. Discussion

While this study utilized conventional bioinformatics tools to successfully clone and characterize the *M. sativa SDH* gene, the integration of AI in future work could profoundly deepen the interpretation of our findings and accelerate functional validation. The specific features of the *M. sativa SDH* sequence uncovered here—such as its acidic pI (5.67), hydrophilic nature, absence of transmembrane domains, and key functional motifs—provide an ideal foundation for AI-driven predictive modeling.

For instance, the amino acid sequence we identified could be used as direct input for deep learning models like AlphaFold [9] to generate a high-accuracy tertiary structure model, moving beyond the preliminary model presented in this study. This could reveal the spatial arrangement of catalytic residues and suggest potential binding mechanisms for substrates or inhibitors. Furthermore, AI algorithms could analyze our phylogenetic results in a broader context, identifying conserved regulatory elements across legumes that control SDH expression under stress conditions [11, 16].

The non-secretory, cytosolic localization predicted for the SDH protein indicates its role in intracellular metabolism. AI models could integrate this subcellular localization data with public transcriptomic datasets to build predictive models of how *M. sativa* SDH expression correlates with drought or pathogen challenge, thereby guiding targeted experimental validation [12]. Finally, the unique sequence motifs we reported could help train convolutional neural networks to identify SDH genes with similar regulatory features in other crops, supporting comparative genomics and precision breeding efforts [8, 13].

In conclusion, the bioinformatic profile of *M. sativa* SDH established in this work provides the essential data layer upon which AI and machine learning can be deployed to transition from *in silico* characterization to *in planta* functional analysis and metabolic engineering.

While this study utilized conventional bioinformatics tools to successfully clone and characterize the *M. sativa SDH* gene, the integration of AI in future work could profoundly deepen the interpretation of our findings and accelerate functional validation. The specific features of the *M. sativa* SDH sequence uncovered here—such as its acidic pI (5.67), hydrophilic nature, absence of transmembrane domains, and key functional motifs—provide an ideal foundation for AI-driven predictive modeling.

For instance, the amino acid sequence we identified could be used as direct input for deep learning models like AlphaFold [9] to generate a high-accuracy tertiary structure model, moving beyond the preliminary model presented in this study. This approach has demonstrated remarkable success in predicting protein structures with atomic-level accuracy, as exemplified by AlphaFold's performance in the CASP14 competition. Such a model could reveal the spatial arrangement of catalytic residues and suggest potential binding mechanisms for substrates or inhibitors, thereby facilitating targeted mutagenesis or inhibitor design. Furthermore, AI algorithms could analyze our phylogenetic results in a broader context, identifying conserved regulatory elements across legumes that control SDH expression under stress conditions [11, 16]. For example, random forest models could be employed to integrate transcriptomic data from public repositories like PhytoMine, enabling the prediction of SDH expression patterns under drought or pathogen challenge and guiding subsequent experimental validation [12].

Additionally, the unique sequence motifs we reported could help train convolutional neural networks to identify SDH genes with similar regulatory features in other crops, supporting comparative genomics and precision breeding efforts [8, 13]. Beyond expression prediction, AI-powered tools such as DeepCRISPR [15] could leverage the gene sequence information to design high-efficiency sgRNAs for CRISPR-based knockout or editing of *M. sativa* SDH, thereby functionally validating its role in stress adaptation or metabolic flux. DeepCRISPR has already been successfully applied in animal and plant systems to improve the efficiency and specificity of gene editing, suggesting its strong potential for use in alfalfa.

In conclusion, the bioinformatic profile of *M. sativa* SDH established in this work provides the essential data layer upon which AI and machine learning can be deployed to transition from *in silico* characterization to *in planta* functional analysis and metabolic engineering.

Acknowledgment

This work was supported by the National Key Research and Development Program of China (Grant No. 2022YFD1802104). The authors would like to thank all members of the College of Life Sciences, Fujian Normal University, for their valuable assistance and collaboration during the preparation of this manuscript.

References

[1] F. X. Niu, Y. P. Du, Y. B. Huang, *et al*. "Recent advances in the production of phenylpropanoic acids and theirderivatives by genetically engineered microorganisms," *Synthetic Biology Journal*, vol. 1, no. 3, pp. 337-357, 2020.

[2] X. Wang, N. Y. Zhu, W. Jiang, S. Y. Si, "Identification of novel anti-tuberculosis lead compound targeting shikimate kinase, " *Acta Pharmaceutica Sinica*, vol. 53, no. 6, pp. 878-886, 2018.

[3] Y. Y. Gong, S. Q. Guo, H. M. Shu, *et al*, "Analysis of Molecular Evolution and Gene Structure of EPSPS Protein in Plant Shikimate Pathway,"*Chinese Bulletin of Botany*, vol. 50, no. 3, pp. 295-309, 2015 .

[4] H. H. Xia and J. Z. Huang, "In Silico Cloning and Analysis of Shikimate Dehydrogenase Gene from Panicum virgatum," *J. Fujian Normal Univ. (Nat. Sci. Ed.)*, vol. 41, no. 3, pp. 97-102, 2025.

[5] F. S. G. Hashemi, M. R. Ismail, M. R. Yusop, M. S. G. Hashemi, M. H. N. Shahraki, H. Rastegari, G. Miah, & F. Aslani. "Intelligent mining of large-scale bio-data: Bioinformatics applications," *Biotechnology & Biotechnological Equipment*, vol. 32, pp. 10-29.2018.

[6] T. Huang, L. Chen, M. Zheng, & J. Song. "Integrated analysis of multiscale large-scale biological data for investigating human disease," *BioMed Research International*, 2015.

[7] H. Poon, C. Quirk, K. Toutanova, & W. Yih. "NLP for precision medicine," *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, pp. 1-2, 2017.

[8] Y. Cai, J. Wang, & L. Deng. "SDN2GO: An integrated deep learning model for protein function prediction," *Frontiers in Bioengineering and Biotechnology*, vol.8, pp. 391. 2020.

[9] A. Senior, R. Evans, J. Jumper, J. Kirkpatrick, L. Sifre, T. Green, C. Qin, A. Žídek, A. W. R. Nelson, A. Bridgland, H. Penedones, S. Petersen, K. Simonyan, S. Crossan, P. Kohli, D. T. Jones, D. Silver, K. Kavukcuoglu, & D. Hassabis. "Improved protein structure prediction using potentials from deep learning," *Nature*, vol. 8, pp. 706-710, 2020.

[10] H. Tong, A. Küken, & Z. Nikoloski. "Integrating molecular markers into metabolic models improves genomic selection for Arabidopsis growth," *Nature Communications*, vol 11, no 2410, 2020.

[11] C. Hill, T. Czauderna, M. Klapperstück, U. Roessner, & F. Schreiber. "Metabolomics, standards, and metabolic modeling for synthetic biology in plants," *Frontiers in Bioengineering and Biotechnology*, vol. 3, no. 167, 2015.

[12] P. Wang, B. M. Moore, S. Uygun, M. D. Lehti-Shiu, C. S. Barry, & S. Shiu. "Optimizing the use of gene expression data to predict plant metabolic pathway memberships," *bioRxiv*. vol.15,no. 204222, 2020.

[13] S. Sun, C. Wang, H. Ding, Q. Zou. "Machine learning and its applications in plant molecular studies," *Briefings in Functional Genomics*, vol. 19, no. 1, pp. 40-48,2019.

[14] H. Deng, Y. Jia, & Y. Zhang. "Protein structure prediction," *International Journal of Modern Physics B*, vol. 31, no. 1741007, pp. 16-19, 2017.

[15] K. Plaimas, J. Mallm, M. Oswald, F. Svara, V. Sourjik, R. Eils, & R. König. "Machine learning based analyses on metabolic networks supports high-throughput knockout screens," *BMC Systems Biology*, vol. 2, no. 67, 2008.

[16] Q. Song, J. Lee, S. Akter, M. Rogers, R. Grene, & S. Li. "Prediction of condition-specific regulatory genes using machine learning," *Nucleic Acids Research*, vol. 48, no. 11, 2020.

AI-Driven Methods for Preservation and Education in Chinese Calligraphy

Jingjian Chun[1], Jiaqi Yi[1]
[1] Education in Development and Management of Education, Srinakharinwirot University.Bangkok Thailand

Abstract—Preserving Chinese calligraphy heritage and enhancing its instruction have become increasingly important in the digital era. Advances in artificial intelligence (AI) provide new tools to analyze, classify, and simulate calligraphic works, thereby enriching traditional teaching and preservation methods. This article surveys how AI techniques—such as image recognition via deep learning and generative modeling—can be applied to Chinese calligraphy. We explore applications including automated style analysis and recognition, AI-assisted tutoring systems, and digital archiving of calligraphic works. These developments demonstrate that AI can significantly improve the transmission, teaching, and conservation of calligraphy.

Index Terms—Chinese calligraphy, Artificial intelligence, Image recognition, Machine learning, Cultural heritage

I. INTRODUCTION

Chinese calligraphy is a revered art form with a rich history dating back thousands of years, and it is considered a valuable world cultural heritage Tens of millions of people around the globe practice or collect calligraphy, and numerous historical works require careful preservation and study Consequently, there is growing interest in computational tools that can support calligraphy research and education. Traditional learning and evaluation of calligraphy have relied on manual apprenticeship and subjective assessment, which are labor-intensive and difficult to scale. In this context, AI offers powerful computational methods to assist calligraphy transmission and preservation.

In particular, machine learning and computer vision allow the digitization of calligraphy images and the automatic extraction of stylistic features. Deep learning models, such as convolutional neural networks, can classify script types or even identify individual calligraphers by recognizing their distinctive stroke patterns This article surveys AI-driven applications in calligraphy analysis, educational technology, and archival preservation. By examining these scenarios, we demonstrate how AI can complement traditional methods and promote cultural continuity.

This work was supported by the Jianlong Innovation and Entrepreneurship Fund under Grant No. BKZZJH202506.Corresponding author: Hongyuan Wang (e-mail: why126@email.cufe.edu.cn).

II. LITERATURE REVIEW

AI-based image recognition techniques have been adapted to the complex patterns of handwritten calligraphy. Modern computer vision models analyze brush strokes, ink intensity, and spatial layout to classify script forms and attribute authorship. Deep CNNs have been employed to extract features unique to each style In practice, a large corpus of labeled calligraphy images (scanned works or handwritten samples) is used to train these models. The network then outputs classification labels or feature embeddings that characterize the calligraphy style.

These images serve as input to convolutional neural networks for style recognition and feature analysis. For instance, an AI model might extract structural features corresponding to the traditional Nine-Palace grid, allowing it to evaluate balance and stroke composition quantitatively.

In experiments, CNN-based classifiers have shown strong performance in style-recognition tasks. For example, a recent study found that a deep learning system correctly identified 960 distinct Chinese characters across five calligraphy scripts with about 95.6% accuracy This suggests that AI can reliably transcribe and classify large character sets, which is crucial for indexing and comparing calligraphy works. Beyond classification, AI methods can also aid in forgery detection and authorship attribution by identifying subtle stylistic anomalies. For instance, if a new piece diverges significantly from learned style patterns, the AI can flag it for expert review.

I I I. THEORETICAL FRAMEWORK

This study is grounded in the concept of "technology-enabled cultural transmission," integrating the Technology Acceptance Model (TAM) with a cognitive-affective-behavioral pathway framework to explain how AI applications in calligraphy education influence student attitudes and behavior.

First, the Technology Acceptance Model (TAM) is employed to interpret how students perceive and accept AI-assisted calligraphy tools, such as augmented reality (AR) stroke simulation systems and haptic-feedback pens. The CHAS scale results show that students rated the perceived usefulness of AI tools at 4.1 on a 5-point scale, indicating a strong belief in the technology's educational potential.

Second, the constructivist learning theory underlines that knowledge is constructed within specific sociocultural contexts. In the context of calligraphy education, AI tools—such as background reconstruction of historical inscriptions and AI-driven stroke decomposition—enhance learners' perceptual understanding of bi yi (the philosophical intent behind brushstrokes). Interview data from instructors indicated that AI-integrated teaching effectively addresses the fragmentation of learning induced by social media consumption.

Additionally, the study introduces a cognitive–affective–behavioral (CAB) pathway model in which:

Cognitive activation is triggered by AI-powered analytical feedback;

Affective resonance is facilitated by multisensory immersive instruction;

Behavioral engagement is reinforced through practice-based outreach programs.

This pathway is supported by the data: the average CHAS score for emotional identification was 4.2, while cognitive understanding lagged at 2.8. Following the integration of AR tools, behavioral intentions increased by 33%. Overall, this study constructs a novel explanatory framework that merges TAM and cultural-cognitive theory to model how AI mediates calligraphy heritage education.

IV. Research Methodology

4.1 Research Design

This study adopts an explanatory sequential mixed methods design that begins with a quantitative analysis of calligraphy students' attitudes toward AI-integrated instruction, followed by qualitative insights from expert interviews and classroom observations. The rationale for this approach lies in the complexity of cultural heritage education, where attitudes, cognition, and behaviors are often shaped not only by instructional content but also by affective and sociocultural dimensions. Through a combination of survey-based statistical modeling and qualitative triangulation, the study aims to capture a holistic view of how AI tools impact the transmission of Chinese calligraphy traditions in higher education.

The design is structured in two primary phases:Phase I (Quantitative): Distribution of a validated Calligraphy Heritage Attitude Scale (CHAS) questionnaire to a stratified sample of university students across five institutions in Shandong Province. Data are analyzed using SPSS to identify patterns, correlations, and significant factors affecting AI acceptance in calligraphy instruction.Phase II (Qualitative): Semi-structured interviews with ten faculty members and program administrators, supplemented by classroom observations and system usage data from AI-enhanced teaching tools such as haptic feedback pens and AR simulation platforms.This mixed-method design enables not only statistical generalization but also contextual understanding—essential for studies situated at the intersection of technology, culture, and pedagogy.

Chinese calligraphy, as an important component of China's excellent traditional culture, has always faced new requirements in protection and education due to the development of the times (Xu Yunchun & Chen Siy i, 2025). In the context of the digital era, the methods of inheriting traditional calligraphy are undergoing profound transformations. Calligraphy protection work has shifted from mere physical preservation to a combination of digital archiving and restoration. Through high-precision scanning and image processing technologies, many precious inscriptions and ink manuscripts have been digitally preserved, laying a solid foundation for academic research and cultural dissemination (Xiao Xiansheng & Chen Long, 2024). These technological methods not only maximize the restoration of the original appearance but also reveal the ink's subtle charm and paper textures through functions such as detail magnification and color correction, offering new perspectives for calligraphy studies.

In the field of calligraphy education, modern technological tools are deeply integrating with traditional teaching methods (Chen Longguo, 2024). The application of multimedia technology has broken through spatial and temporal limitations in calligraphy instruction. Students can observe the entire process of masters writing through high-definition videos, closely studying the turns of the brush and the variations of ink tones. The establishment of digital character databases provides learners with abundant models for copying. Calligraphy works in different scripts and styles can be presented in standardized formats, enabling effective comparative study (Guo Cheng, 2025). Especially at the elementary education stage, the use of multimedia courseware and interactive teaching software makes stroke practice, which was once relatively monotonous, more vivid and engaging, thereby effectively enhancing students' interest. Teachers can also record each student's practice process via digital platforms, conduct longitudinal comparisons, and provide personalized guidance to achieve differentiated instruction (Li Xiaoyan, 2025).

It is worth noting that the application of modern technologies in calligraphy teaching extends beyond skill training to the realm of cultural inheritance (Siqingga, 2025). By establishing digital calligraphy museums, students can virtually visit treasured original works by masters from various regions and learn about the historical stories and cultural connotations behind the pieces. Some teaching software also incorporates knowledge of calligraphy history and philology, enabling learners to gain a deeper understanding of the cultural essence of calligraphy while mastering writing skills. This comprehensive digital learning experience helps cultivate students' holistic understanding of traditional culture and their aesthetic abilities.

However, a sober awareness is needed in advancing the digitalization of calligraphy (Chen Zhenlian, 2017). Technology is ultimately an auxiliary tool and cannot replace the humanistic care and emotional interaction embodied in the

traditional teacher-apprentice mode of instruction. The essence of calligraphy lies in the spirit and personality conveyed between brush and ink, which no technology can fully replicate (Xing Tiantian & Ding Shaoshuai, 2024). Therefore, while applying modern technologies, greater emphasis should also be placed on the inheritance and innovation of traditional teaching methods so that the two can complement each other. Teachers play a key role in this process: they need to continuously enhance their own competencies, mastering new technologies while deeply understanding the essence of calligraphy art, in order to better guide students to appreciate the beauty of calligraphy (Wang Peng, 2025).

Currently, calligraphy education is moving toward a stronger emphasis on cultural connotation and aesthetic cultivation (Xu Shukun, 2025). Many educational institutions have begun experimenting with integrating calligraphy with other art forms to carry out interdisciplinary teaching practices. For example, combining calligraphy with classical literature, traditional music, and Chinese painting allows students to experience the profoundness of Chinese culture through multiple art forms. This comprehensive approach not only enriches the content and methods of calligraphy education but also better aligns with the requirements of contemporary quality-oriented education (Yu Qiang, 2025).

In conclusion, modern technologies provide new possibilities for the protection and education of calligraphy, but the ultimate goal is always to better transmit and develop this ancient art. On the premise of maintaining the essential characteristics of calligraphy, the rational application of technology and the continuous innovation of teaching methods can bring new vitality to the art in the new era. This process requires joint efforts from educators, cultural scholars, and technology experts, as well as broad participation and support from all sectors of society. Only in this way can the contemporary inheritance and innovative development of calligraphy truly be realized, allowing this ancient art to flourish brilliantly under modern conditions.

V. CONCLUSION AND RECOMMENDATIONS

5.1 Summary of Findings

Based on integrated data analysis, the study draws the following conclusions:AI Technologies Enhance Technical Mastery but Require Pedagogical Framing Tools such as AR and AI-based stroke evaluation significantly improved learners' execution of calligraphic forms. However, without instructional framing rooted in cultural context, students struggled to grasp philosophical meanings, affirming the need for integrated mentorship.

Cultural Identification Is a Strong Predictor of Engagement Emotional attachment to calligraphy traditions—often cultivated through immersive, multisensory instruction—proved more influential than technical knowledge in predicting students' behavioral intentions toward preservation.

Mentorship + AI = A Scalable but Human-Centered Model The 1+N mentorship model (one ICH bearer + interdisciplinary AI mentors) successfully preserved the authenticity of oral tradition while leveraging AI's scalability. This model yielded an 85% accuracy rate in historical context reconstructions and a 40% gain in stroke reproduction precision.Institutional and Policy-Level Gaps Persist Only 22% of surveyed institutions had dedicated cultural heritage educators, and fewer than 30% had allocated budgets for AI teaching tools. Cultural compression, as amplified by decontextualized digital content, continues to threaten holistic heritage transmission.

5.2 Policy and Educational Recommendations

To improve AI-enabled calligraphy education, we propose the following strategies:

Curriculum Innovation:Develop AI-integrated calligraphy modules that include AR-based inscription simulations, VR galleries of ancient scrolls, and real-time stroke feedback systems.Faculty Development:Organize interdisciplinary training workshops where calligraphy instructors collaborate with AI developers to co-create pedagogical content and assessment rubrics.Policy Intervention:Mandate inclusion of ICH education in university curricula. Allocate a minimum of 3% of cultural budgets to AI infrastructure supporting traditional arts.Data and Infrastructure Support:Scale up the Yellow River Basin Calligraphy DNA Project, using blockchain to archive high-resolution (1200 dpi) stele scans and crowdsource annotations for AI training.Public–Private Partnerships:Collaborate with cultural enterprises to commercialize digital calligraphy products and establish mentorship-based artist residencies that integrate AI tools into traditional apprenticeship pathways.

5.3 Contributions and Future Directions

Theoretical Contribution This study is the first to embed traditional apprenticeship models into the TPACK framework (Technological Pedagogical Content Knowledge), offering a novel interdisciplinary approach to cultural education in the digital era.Practical Contribution It provides an empirically grounded implementation strategy for UNESCO SDG 11.4, which focuses on safeguarding cultural heritage, demonstrating measurable improvements in learner engagement, accuracy, and long-term retention.Limitations and Future Work The research sample focused primarily on students enrolled in calligraphy majors. Future studies should expand to non-art disciplines to evaluate transferability. Additionally, institutional resistance to educational innovation warrants deeper exploration through organizational change theories and comparative policy analysis.

Reference

1．Aboagye, S. (2023). The impact Confucius on education and culture in China and Ghana. ResearchGate. https://doi.org/10.13140/RG.2.2.11219.66082

2．ARAL, A. (2018). Intangible cultural heritage and education: critical examination of the education in periodic reports. Milli Folklor(120).

3．Bandura, A. (2018). Toward a Psychology of Human Agency: Pathways and Reflections. Perspect Psychol Sci, 13(2), 130-136. https://doi.org/10.1177/1745691617699280

4．Barghi, R., Hamzah, A., & Rasoolimanesh, S. M. (2020). To what extent Iranian primary school textbooks mirror the philosophy of heritage education? Journal of Cultural Heritage Management and Sustainable Development, 11(1), 58-77. https://doi.org/10.1108/jchmsd-12-2018-0087
5．Gaikwad, Y. Z. S. S. (2025). Analysis on the Integration of Henan Traditional Art Intangible Cultural Heritage into Public Art Education in Local Colleges and Universities. Journal of Information Systems Engineering and Management.
6．Gao, Q., & Sawadee, Y. (2024). Design of Calligraphy Teaching Materials from the Perspective of Chinese Preschool Education-Anshan City Case Study. In International Journal of Sociologies and Anthropologies Science Reviews (pp. 91–102).
7．Guo, Y. (2024). Potentials of arts education initiatives for promoting emotional wellbeing of Chinese university students. Front Psychol, 15, 1349370. https://doi.org/10.3389/fpsyg.2024.1349370
8．Huang, X., & Qiao, C. (2024). The Effects and Learners' Perceptions of Cluster Analysis-Based Peer Assessment for Chinese Calligraphy Classes. SAGE Open, 14(2). https://doi.org/10.1177/21582440241255846
9．Hutson, J., Weber, J., & Russo, A. (2023). Digital Twins and Cultural Heritage Preservation: A Case Study of Best Practices and Reproducibility in Chiesa dei SS Apostoli e Biagio. Art and Design Review, 11(01), 15-41. https://doi.org/10.4236/adr.2023.111003
10．Lee, L. Y. S. (2022). Community of practice: the making of knowledge dynamic in intangible cultural heritage. Consumer Behavior in Tourism and Hospitality, 17(3), 338-350. https://doi.org/10.1108/cbth-11-2021-0278
11．Li, H. (2023). New vision on calligraphy general education of college students from the perspective of anthropology. Advances in Educational Technology and Psychology, 7(12). https://doi.org/10.23977/aetp.2023.071206
12．Li, L., & Tang, Y. (2023). Towards the Contemporary Conservation of Cultural Heritages: An Overview of Their Conservation History. Heritage, 7(1), 175-192. https://doi.org/10.3390/heritage7010009
13．Li, R., Jia, X., Zhou, C., & Zhang, J. (2022). Reconfiguration of the brain during aesthetic experience on Chinese calligraphy—Using brain complex networks. Visual Informatics, 6(1), 35-46. https://doi.org/10.1016/j.visinf.2022.02.002
14．Li, Y. (2023). Exploration of Calligraphy Education among University Students in Shanxi Province of China. Frontiers in Educational Research, 6(21). https://doi.org/10.25236/fer.2023.062129
15．Li, Y., & Zhang, W. (2021). Interpretation of the Convention for the Safeguarding of Intangible Cultural Heritage. China Intangible Cultural Heritage, 6, 109–113.
16．Li, Y. Z., W. (2022). Current status and challenges of calligraphy education in Chinese universities. In Journal of Chinese Calligraphy Studies.
17．López-Fernández, J. A., Medina, S., López, M. J., & García-Morís, R. (2021). Perceptions of heritage among students of early childhood and primary education. Sustainability, 13(19), 10636.
18．Minerva, R., Lee, G. M., & Crespi, N. (2020). Digital Twin in the IoT Context: A Survey on Technical Features, Scenarios, and Architectural Models. Proceedings of the IEEE, 108(10), 1785-1824. https://doi.org/10.1109/jproc.2020.2998530
19．Morch, V. M. V. H. K. B. B. B. G. J. M. M. N. P. R. R. R. V.-M. (2022). The Role of Robotics in Achieving the United Nations Sustainable DevelopmentGoals—The Experts' Meeting at the 2021 IEEE/RSJ IROS Workshop [Industry Activities]. https://doi.org/10.1109/MRA.2022.3143409
20．Mukherjee, Y., & Palit, S. (2022). Digitalisation and Revitalisation of Cultural Heritage through Information Technology. In Digitalization of Culture Through Technology (pp. 44-49). https://doi.org/10.4324/9781003332183-8
21．Samuel J, K. (2024). The Impact of Creative Arts on Student Engagement and Learning. Research Invention Journal of Research in Education, 4(1), 1-5. https://doi.org/10.59298/rijre/2024/4115
22．Shen, W., Shi, J., Meng, Q., Chen, X., Liu, Y., Cheng, K., & Liu, W. (2022). Influences of Environmental Regulations on Industrial Green Technology Innovation Efficiency in China. Sustainability, 14(8). https://doi.org/10.3390/su14084717
23．Skalkos, D., Kosma, I. S., Chasioti, E., Skendi, A., Papageorgiou, M., & Guiné, R. P. F. (2021). Consumers' Attitude and Perception toward Traditional Foods of Northwest Greece during the COVID-19 Pandemic. Applied Sciences, 11(9). https://doi.org/10.3390/app11094080
24．Wang, P., & Wu, T. (2022). A Study on the Holographic Value of Calligraphy Inheritance—Taking Dunhuang Posthumous Paper of Wei-Jin Period as an Example The 2021 Summit of the International Society for the Study of Information,
25．Wang, X. C., H. (2020). Integrating augmented reality into traditional calligraphy pedagogy: A case study in Shandong Province. In Educational Technology Research and Development.
26．Xia, Y., Deng, Y., Tao, X., Zhang, S., & Wang, C. (2024). Digital art exhibitions and psychological well-being in Chinese Generation Z: An analysis based on the S-O-R framework. Humanities and Social Sciences Communications, 11(1). https://doi.org/10.1057/s41599-024-02718-x
27．XingJia, T., PengChang, Z., ZongBen, X., & BingLiang, H. (2022). Calligraphy and Painting Identification 3D-CNN Model Based on Hyperspectral Image MNF Dimensionality Reduction. Comput Intell Neurosci, 2022, 1418814. https://doi.org/10.1155/2022/1418814
28．Chen, L. G. (2024). Research on the development and application of calligraphy education in the AI era. Calligraphy Education, (3), 95–97.

29．Chen, Z. L. (2017). How can calligraphy “transform gracefully” in the age of artificial intelligence? Art Observation, (27), 43.

30．Guo, C. (2025). Analysis of the application of artificial intelligence technology in calligraphy teaching. Calligraphy Education, 12.

31．Li, X. Y. (2025). Practice and reflection on intelligent technology empowering calligraphy teaching. Digital Teaching in Primary and Secondary Schools, (2), 34–37.

32．Siqingga. (2025). Development of regional culture-oriented calligraphy curriculum empowered by artificial intelligence. Research on Ethnic Education, (1), 88–92.

33．Wang, P. (2025). Role transformation and capacity enhancement of calligraphy teachers in the digital-intelligence era. Chinese Journal of Education, (4), 56–60.

34．Xing, T. T., & Ding, S. S. (2024). The impact of artificial intelligence on traditional calligraphy in the new era. Art & Technology, (2), 21.

35．Xu, S. K. (2025). Path exploration of innovation in calligraphy education driven by artificial intelligence. China Tourism News, p.004.

36．Xu, Y. C., & Chen, S. Y. (2025). Transformation and persistence of calligraphy education in primary and secondary schools in the AI era. Chinese Art Education, (3), 14–22.

37．Xiao, X. S., & Chen, L. (2024). Exploration of calligraphy teaching paths in art colleges under the background of artificial intelligence. Art Education, (5), 76–78.

38．Yu, Q. (2025). The significance of artificial intelligence technology to primary and secondary school calligraphy education. Calligraphy Education, 10.