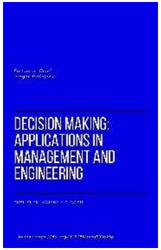
SCIENTIFIC OASIS

Decision Making: Applications in Management and Engineering

Journal homepage: www.dmame-journal.org
ISSN: 2560-6018, eISSN: 2620-0104



An Intelligent Decision-Support Framework Using Mobile and Virtual Reality Technologies for Optimising Intangible Cultural Heritage Management

Chen Mao*

1 Department of Art Design, Xiamen Nanyang Vocational College, Xiamen 361102, China. Email: maochen83@sina.com

ARTICLE INFO

Article history:
Received 15 July 2025
Received in revised form 29 October 2025
Accepted 09 December 2025
Available online 30 December 2025

Keywords:
Intangible Cultural Heritage (ICH), Decision-Support System, Multimodal Data Fusion, Adaptive Convolutional Neural Network (ACNN), Intelligent Long Short-Term Memory (ILSTM), Virtual Reality (VR)

ABSTRACT

Effective administration of intangible cultural heritage (ICH) increasingly faces challenges arising from intricate prioritization demands, constrained resources, and sustainability obligations. Conventional approaches to ICH management are predominantly manual and rely heavily on expert judgement, limiting their scalability and their ability to integrate data across multiple modalities. This study introduces an intelligent decision-support framework that incorporates interactive mobile interfaces, virtual reality (VR), and a novel Adaptive Convolutional Synthesized Intelligent Memory Network (ACSIMN) to facilitate precise ICH prioritization and optimise resource distribution through adaptive learning of multimodal features and cross-modal reasoning. Data are gathered from diverse sources, including images and videos of cultural artefacts, audio recordings of traditional music and oral histories, and textual materials detailing historical and cultural contexts. Modality-specific pre-processing procedures are applied, such as resizing visual inputs for images and videos, generating spectrograms from audio signals, and tokenising and cleansing textual information. Feature extraction is conducted using Scale-Invariant Feature Transform (SIFT) for visual inputs, Mel-Frequency Cepstral Coefficients (MFCCs) for audio data, and Bidirectional Encoder Representations from Transformers (BERT) for text. These features are then integrated into a unified latent space, enabling coherent multimodal correlation and informed joint decision-making. The principal innovation resides in the ACSIMN model, which harmonises adaptive convolutional neural network (ACNN) learning for visual interpretation with intelligent long short-term memory (ILSTM) for temporal and semantic sequence modelling, complemented by VR-enabled visualisation to improve situational comprehension. Implemented in Python, the proposed framework attains an accuracy of 98.4%, supporting both accurate recognition and real-time prioritization of cultural assets, while demonstrating potential for sustainable, scalable, and data-driven preservation of ICH.

1. Introduction

ICH encompasses the evolving cultural practices of contemporary communities, including

[*] *Corresponding author.*
E-mail address: maochen83@sina.com

https://doi.org/10.31081/dmame8220251631

traditional rituals, performing arts, and oral traditions, maintained across generations. It is crucial for preserving intergenerational knowledge, social identity, and cultural diversity [20]. Managing ICH is inherently complex, involving priority-setting, documentation, and sustainability considerations. With finite resources, the presence of mobile cultural systems, and regional disparities, structured and informed management strategies are essential for heritage preservation [1]. Digital VR technologies enhance documentation and dissemination of cultural heritage, making it more interactive and widely accessible. They enable community participation in conserving intangible cultural expressions, overcoming geographical barriers [11]. Interactive VR and mobile technologies serve as versatile tools for real-time recording, engagement, and involvement of communities, supporting inclusive and participatory management approaches [12].

The integration of VR and augmented reality as immersive digital experiences enriches the understanding of cultural practices within real-world contexts, fostering awareness, evaluation, and appreciation of ICH [10]. Data-driven methodologies facilitate the analysis of large multicultural datasets, providing objective assessments, revealing trends, and supporting effective cultural management through systematic interpretation of visual, auditory, and textual data [12]. Cultural expressions are inherently multimodal, typically combining visual, auditory, and narrative elements. These sensory dimensions contribute to a more comprehensive and realistic representation of cultural phenomena [3]. Heritage policymakers and managers increasingly adopt decision-support systems to evaluate cultural significance, accessibility, and sustainability, thereby improving transparency, accountability, and consistency in heritage management [6].

Despite these advances, a cohesive framework that integrates interactivity, immersive visualisation, and analytical support for ICH management remains largely absent [2]. There is growing demand for collaborative systems that seamlessly connect digital and data-driven technologies to enable more sustainable and informed heritage practices [13]. The workflow for employing mobile and VR technologies to optimise ICH management is illustrated in Figure 1.

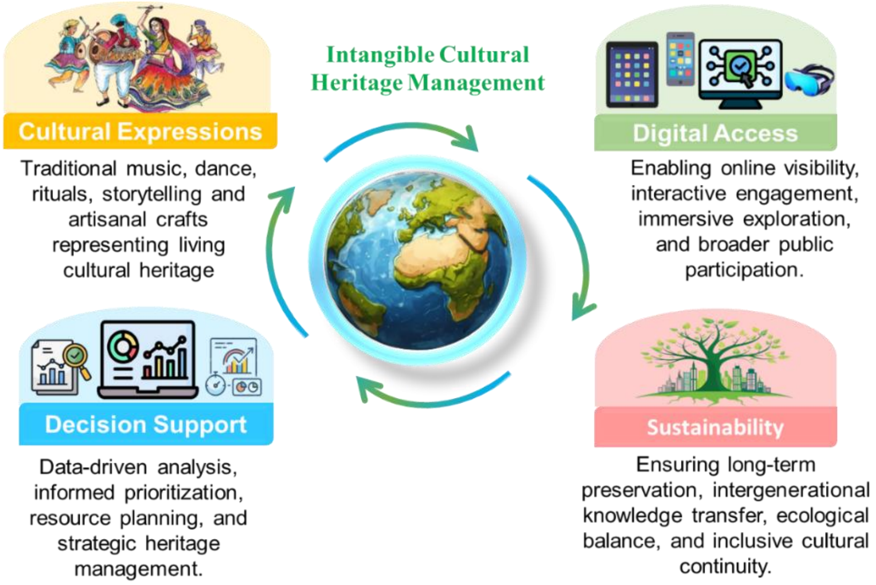

Intangible Cultural Heritage Management
Cultural Expressions
Traditional music, dance, rituals, storytelling and artisanal crafts representing living cultural heritage
Digital Access
Enabling online visibility, interactive engagement, immersive exploration, and broader public participation.
Decision Support
Data-driven analysis, informed prioritization, resource planning, and strategic heritage management.
Sustainability
Ensuring long-term preservation, intergenerational knowledge transfer, ecological balance, and inclusive cultural continuity.

Fig.1: Process of Mobile and VR Technologies in Enhancing ICH Management

1.1 Research Aim

The study seeks to develop an intelligent decision-support system for ICH by integrating ACNN with ILSTM, further enhanced through the fusion of multimodal features. The designed framework is intended to accurately prioritise and analyse diverse cultural assets, including images, videos, audio recordings, and textual data, while improving recognition precision, enabling cross-modal reasoning, and providing resilience for applications in mobile and VR-based heritage management.

1.2 Research Organization

Section 1 outlines the rationale for adopting intelligent approaches to ICH management, while Section 2 reviews existing techniques for ICH preservation. Section 3 introduces a decision-support framework based on the ACSIMN architecture. Section 4 details the experimental design, performance assessment, and obtained results. Finally, Section 5 explores prospects for sustainable and scalable heritage management.

1.3 Related Works

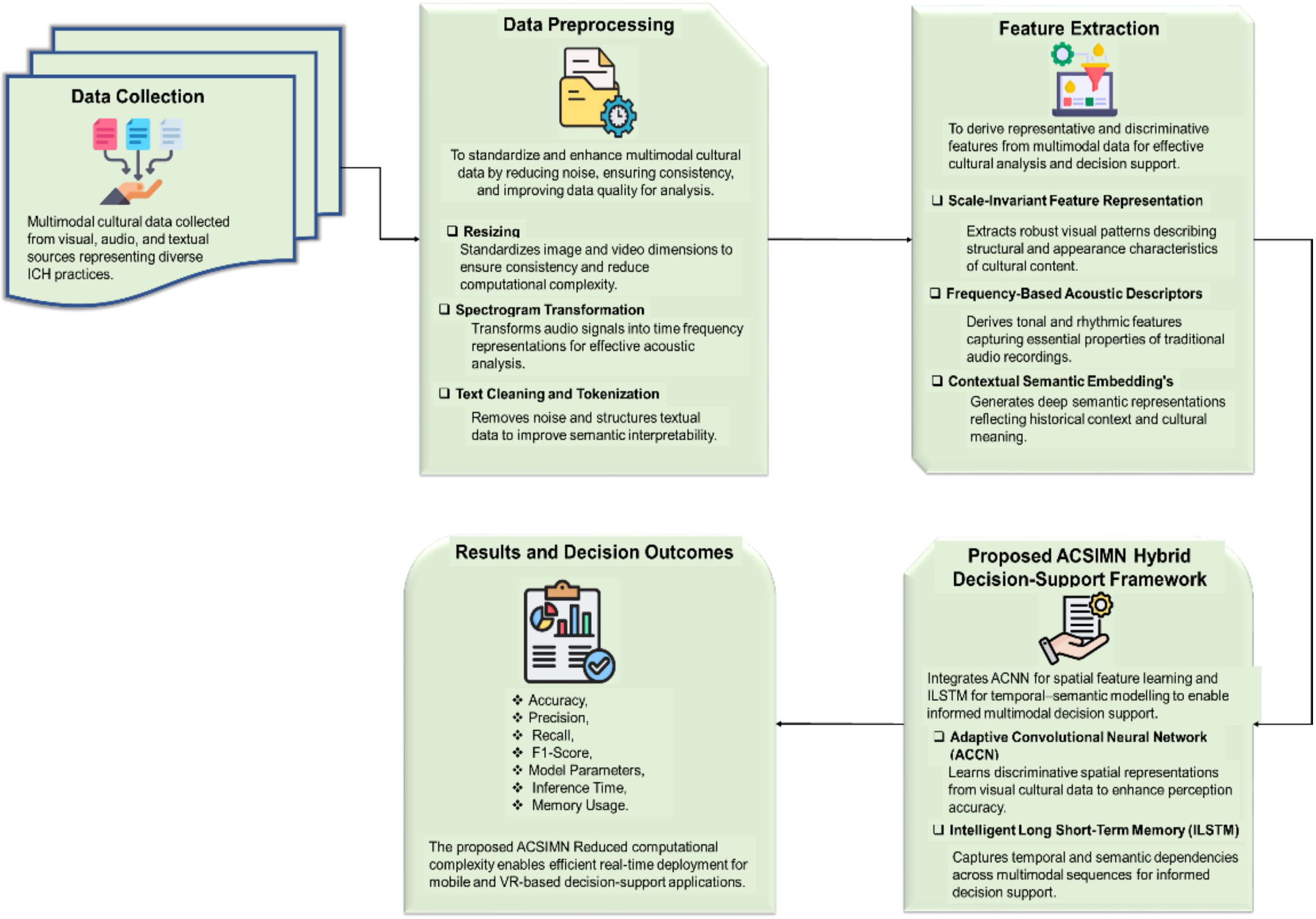
Table 1 provides a summary of research on mobile and VR technologies, as well as multimodal learning approaches applied to ICH management. It highlights study objectives, employed methodologies, key findings, and identified limitations, emphasises existing research gaps, and underscores the need for an integrated intelligent decision-support system.

Table 1
Comparative Analysis of Existing Studies on Intelligent, Mobile, and VR Technologies for ICH Management

Reference	Method	Objective	Key Results	Limitations
Galani and Vosinakis [9]	Mobile augmented reality with three-dimensional (3D) overlays	Enhance user engagement and appreciation of cultural heritage	Improved learning outcomes and user experience; system reported as intuitive and user-friendly	Small participant sample; limited comparison with digital-only platforms
Di Giulio et al. [5]	Vision Transformer combined with linear discriminant analysis (LDA)	Facilitate ICH management via heritage value assessment	Model achieved 0.889 accuracy	Needs recalibration for different cultural contexts
Fu et al. [8]	Deep learning (DL) and Natural Language Processing (NLP) systems	Support efficient documentation of ICH	Cultural accuracy of 92% with increased community involvement	Ethical considerations and sustainability issues
Wu et al. [19]	Lightweight CNN + RNN with optimisation	Automate visual storytelling for ICH preservation	High reliability, exceeding 98% accuracy	Dataset diversity was limited
Ziku et al. [24]	Affective Event Theory model	Assess the effect of digital experiences on ICH dissemination	Visual and auditory elements improved dissemination behaviour	Dependent on self-reported survey data
Xie [20]	VR-based digital mobile display technology	Enhance interaction and realism in ICH exhibitions	Visual experience improved by over 25%	Implementation costs were high
Zhang et al. [21]	AI, VR, digital twins, and blockchain	Evaluate effectiveness of heritage conservation (HCE)	Blockchain enabled authentic digital experiences for HCE	Potential bias in simulations
Cui [4]	3D scanning and digital photogrammetry	Preserve and visualise ICH	Digital resource development improved	User engagement remained limited
Zhang and Li [23]	Immersive VR with gamified reconstruction	Systematic reconstruction and long-term preservation of ICH	Enhanced user experience, cultural continuity, and preservation outcomes	Technical complexity and long-term sustainability concerns
Sun et al. [17]	Multimodal framework	Enable scalable documentation and indexing of ICH	Achieved 94.8% accuracy with better information transmission	Requires expert involvement
Shang [15]	ALS-optimized MKSVM-LSTM	Predict ethnic assessment of ICH artworks	Achieved 97% accuracy, outperforming conventional models	Dataset size limited; potential feature bias

2 Methodology

ACSIMN is a methodological framework that integrates ACNN, ILSTM, and multimodal feature fusion to enable comprehensive analysis of ICH. Through pre-processing of diverse data types—including images, videos, audio, and textual content—the approach improves data quality and uniformity. ACNN is employed to extract spatial features from visual inputs, while ILSTM models temporal and semantic patterns in audio and textual data, facilitating cross-modal reasoning within a unified latent representation. The framework also incorporates VR-based visualisation to enhance situational comprehension, as depicted in the workflow diagram in Figure 2.



Multimodal cultural data collected from visual, audio, and textual sources representing diverse ICH practices.

To standardize and enhance multimodal cultural data by reducing noise, ensuring consistency, and improving data quality for analysis.

❑ **Resizing**
Standardizes image and video dimensions to ensure consistency and reduce computational complexity.

❑ **Spectrogram Transformation**
Transforms audio signals into time frequency representations for effective acoustic analysis.

❑ **Text Cleaning and Tokenization**
Removes noise and structures textual data to improve semantic interpretability.

To derive representative and discriminative features from multimodal data for effective cultural analysis and decision support.

❑ **Scale-Invariant Feature Representation**
Extracts robust visual patterns describing structural and appearance characteristics of cultural content.

❑ **Frequency-Based Acoustic Descriptors**
Derives tonal and rhythmic features capturing essential properties of traditional audio recordings.

❑ **Contextual Semantic Embedding's**
Generates deep semantic representations reflecting historical context and cultural meaning.

Integrates ACNN for spatial feature learning and ILSTM for temporal–semantic modelling to enable informed multimodal decision support.

❑ **Adaptive Convolutional Neural Network (ACCN)**
Learns discriminative spatial representations from visual cultural data to enhance perception accuracy.

❑ **Intelligent Long Short-Term Memory (ILSTM)**
Captures temporal and semantic dependencies across multimodal sequences for informed decision support.

❖ Accuracy,
❖ Precision,
❖ Recall,
❖ F1-Score,
❖ Model Parameters,
❖ Inference Time,
❖ Memory Usage.

The proposed ACSIMN Reduced computational complexity enables efficient real-time deployment for mobile and VR-based decision-support applications.

Fig.2: Proposed Framework for Cultural Heritage Analysis

2.1 Data Acquisition

Cultural Heritage Management Data (https://www.kaggle.com/datasets/colabsss/cultural-heritage-management-dataset?resource=download&select=ich_video_dataset.csv) is designed to support intelligent decision systems for the managing and prioritization of ICH in mobile and VR environments. The system captures and integrates cultural assets through multimodal data, comprising 2,000 audio recordings, 2,000 videos, and 2,000 textual entries corresponding to various ICH items. It draws from ten national repositories of images and other cultural expressions, providing details on cultural context, geographic location, temporal relevance, documentation, community participation, and sustainability metrics. Each cultural resource is evaluated according to preservation priority and resource allocation requirements. The dataset is further divided into training (80%) and testing (20%) subsets to enable evaluation of intelligent frameworks for recognition, prioritization, and management of ICH.

2.2 Pre-Processing

To ensure visual consistency and reduce source variability, images and videos of cultural artefacts are first standardised to uniform resolutions. Audio recordings of traditional music and oral narratives are converted into spectrograms, with temporal frequency patterns extracted to enable effective analysis. Textual documents are tokenised into words, phrases, or symbols, creating structured inputs that preserve cultural context and support accurate recognition and prioritisation of ICH assets.

2.2.1 Image Resizing for Consistent Input to Achieve Accurate ICH Prioritization

To ensure reliable analytical outcomes for cultural artefacts, images and videos are down-sampled to mitigate variations in resolution, scale, and aspect ratio. Visual data are resized to a fixed resolution using bi-cubic interpolation, preserving the original aspect ratio and retaining essential content. This standardisation guarantees uniformity, maintains visual integrity, and supports effective management of ICH assets. Figure 3 illustrates the process with (a) original and (b) resized images (256×256), demonstrating how pre-processing aligns input dimensions without altering visual details, thereby enabling consistent feature extraction across multimodal heritage analyses.

Fig.3: Representations of (a) Original and (b) Resized Cultural Artefact Images

2.2.2 Spectrograms for Temporal Analysis of Traditional Audio in ICH

Audio recordings of traditional music and oral narratives contain rich temporal and frequency information that is essential for analysing cultural heritage. These recordings are transformed into spectrograms to enable accurate recognition and prioritisation of ICH. The spectrograms capture variations in time, frequency, rhythm, pauses, and speech rate, supporting systematic audio processing, seamless multimodal integration, and informed decision-making for the effective preservation of cultural assets.

2.2.3 Tokenization for Structured Text Analysis in ICH

The analysis of ICH textual data begins with tokenization, which involves segmenting documents into smaller units, or tokens, such as words, phrases, or symbols, to facilitate effective feature extraction. This process systematises text processing, producing structured inputs suitable for subsequent embedding or sequential modelling. Common tokenization strategies include division by spaces, punctuation, or line breaks. Advanced tokenization preserves significant historical and cultural patterns, reduces ambiguity, and enhances the quality of textual representations, thereby improving recognition, classification, and prioritisation of ICH items.

2.3 Feature Extraction

SIFT is applied to images and video data to extract scale-invariant visual features, capturing both structural and appearance characteristics of cultural artefacts. MFCC is utilised to analyse audio spectrograms, deriving frequency-based attributes that encode tonal, rhythmic, and temporal information. BERT is employed to generate contextual semantic embeddings of textual data, representing historical and cultural meaning, thereby enabling effective integration of ICH assets across modalities and supporting more informed prioritisation.

2.3.1 Feature Extraction Using SIFT for Visual Analysis of ICH Artefacts

Visual representations of cultural elements and traditional practices exhibit substantial variations in scale, orientation, and viewing conditions due to differing acquisition environments. To address this variability, SIFT is employed to characterise images through local key points that remain invariant to changes in scale and rotation, making it particularly suitable for heterogeneous ICH image datasets. Key point detection is performed within a scale-space using the Difference of Gaussian (DoG) method. The Gaussian scale-space representation is defined in Equation (1):

$$H(i,j,\sigma) = Y(i,j,\sigma) * K(i,j) \tag{1}$$

Where $H(i,j)$ represents the input cultural image, $Y(i,j,\sigma)$ is the Gaussian kernel, σ is the standard deviation σ, $K(i,j,\sigma)$ denotes the scale-space image, and $*$ indicates convolution. Stable key points are identified by locating extrema across neighbouring scales. Each significant point is assigned an orientation based on local image gradients, ensuring invariance to rotation. The gradient magnitude is computed using Equation (2):

$$p(i,j) = \sqrt{(H(i+1,j) - H(i-1,j))^2 + (H(i,j+1) - H(i,j-1))^2} \tag{2}$$

$p(i,j)$ denotes the gradient magnitude at pixel location i,j. $H(i+1,j$ and $H(i-1,j)$ are the values of H at the neighbouring pixels above and below (i,j). These gradients capture structural characteristics of cultural artefacts, allowing their visual representation to remain consistent and supporting reliable prioritisation and management of ICH assets.

Figure 4 illustrates the processing of cultural artefact images and their features using the SIFT algorithm, highlighting key points, edges, and textures as critical components in heritage analysis through multimodal representation.

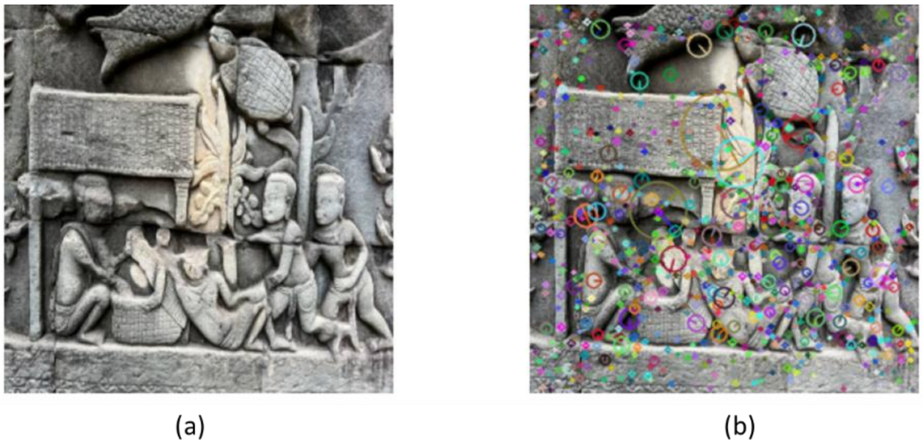


Fig.4: Representation of (a) Pre-Processing and (b) Feature Extraction of Cultural Artefacts

2.3.2 MFCC Feature Extraction for Traditional Audio Analysis in ICH

Audio recordings of traditional music and oral narratives contain rich perceptual information, making them critical for ICH analysis. To extract these features, MFCCs are employed, as they model human auditory perception and emphasise frequency-based components that are perceptually significant. Equation (3) defines the nonlinear connection between the actual frequency e (in Hz) and the Mel frequency e_{Mel}.

$$e_{\text{Mel}} = 2595 \times \log_{10}\left(1+\frac{e}{700}\right) \tag{3}$$

where e_{Mel} represents the Mel-scaled frequency aligned with human hearing sensitivity. This process is instigated by dividing the discrete audio signal into low overlying edges $w_j(m)$, where j denotes the frame index and $m$denotes the sample index within each frame.

Equation (4), representing the Fast Fourier Transform (FFT), is applied to convert each individual frame into its corresponding frequency spectrum.

$$W(j,l) = \sum_{m=0}^{M-1} w_j\,(m) X_M^{lm}, l = 0,1,\ldots,M-1 \tag{4}$$

Here, $W(j,l)$ is the frequency-domain representation of frame $j$at frequency bin l, $M$is the total number of samples per frame, and X_M^{lm} denotes the FFT basis function. The associated power spectrum is then calculated by Equation (5).

$$F(j,l) = \mid W(j,l) \mid^2 \tag{5}$$

After that, the power spectrum is run through a bank of N Mel-scaled filters to get the Mel filter bank energies, which are determined using Equation (6).

$$T(j,k) = \sum_{l=0}^{M-1} F(j,l)\ G_K(l) \tag{6}$$

Where $T(j,k)$ represents the findings of the k-th Mel filter $F(j,l)$ for frame j, and $G_k(l)$ is the weighting function of the n-th Mel filter. Finally, the logarithm of the Mel filter energies is transformed using a Discrete Cosine Transform (DCT) to derive the MFCCs, as expressed in Equation (7).

$$MFCC(j,i) = \sqrt{\frac{2}{N}} \sum_{n=0}^{K-1} \log\left[T(j,k)\right] \cos\left(\frac{\pi i(n-0.5)}{N}\right), m = 1,2,\ldots,K \tag{7}$$

$MFCC(j,i)$ denotes the $i-th$ spectral coefficient of frame j, $logzT(j,k)]$ logarithmic compression, cos(·) is a cosine basis. π is a mathematical constant, n denotes coefficient order N denotes the total of Mel filters, and $K$is a number of retained MFCC coefficients. MFCC representations effectively capture the key characteristics of traditional audio, enabling their seamless integration with visual and textual data for consistent analysis and prioritisation of ICH assets.

2.3.3 BERT for Semantic Understanding and Cultural Context Representation in Textual Data

BERT is employed to capture rich contextual features from textual materials associated with ICH, such as oral narratives, traditional stories, and ritual descriptions. Its bidirectional processing enables the model to learn semantic and syntactic patterns, generating embeddings that encode both meaning and context within ICH content. These deep contextual representations are then integrated into multimodal analyses, allowing textual data to be accurately classified, aligned, and fused with audio and visual modalities to support effective documentation and preservation of ICH.

2.4 Multimodal Data Fusion

Multimodal data fusion entails integrating visual, audio, and textual information into a unified latent representation that preserves cross-modal semantic relationships. Feature representations obtained from SIFT, MFCC, and BERT are standardised, aligned, and jointly encoded to ensure consistent correlations across heterogeneous modalities. This coherent representation supports high-quality collaborative inference, enhancing both the accuracy and consistency of decisions regarding the prioritisation and management of ICH.

2.5 ACSIMN for Intelligent ICH Decision Support

ACSIMN integrates complementary learning mechanisms to enable intelligent and sustainable management of ICH. The ACNN component extracts spatial patterns from visual cultural data, providing robust representations under diverse conditions. The ILSTM module models temporal and semantic relationships in audio and textual narratives, preserving the sequential cultural context. By unifying these components within a single hybrid framework, ACSIMN supports coherent multimodal reasoning, facilitating informed decision-making and prioritisation in the preservation of ICH.

2.5.1 ACNN for Intelligent Analysis and Categorisation of ICH

The ACNN is a supervised deep learning model employed for accurate identification and classification of ICH elements. It is utilised to extract discriminative visual features from images and videos of traditional artefacts, crafts, and cultural performances, supporting proper digitalisation and long-term heritage preservation. Figure 5 illustrates the ACNN architecture, highlighting hierarchical feature extraction, max pooling, fully connected layers, and SoftMax-based classification, which collectively enable effective learning and discrimination of multimodal cultural representations.

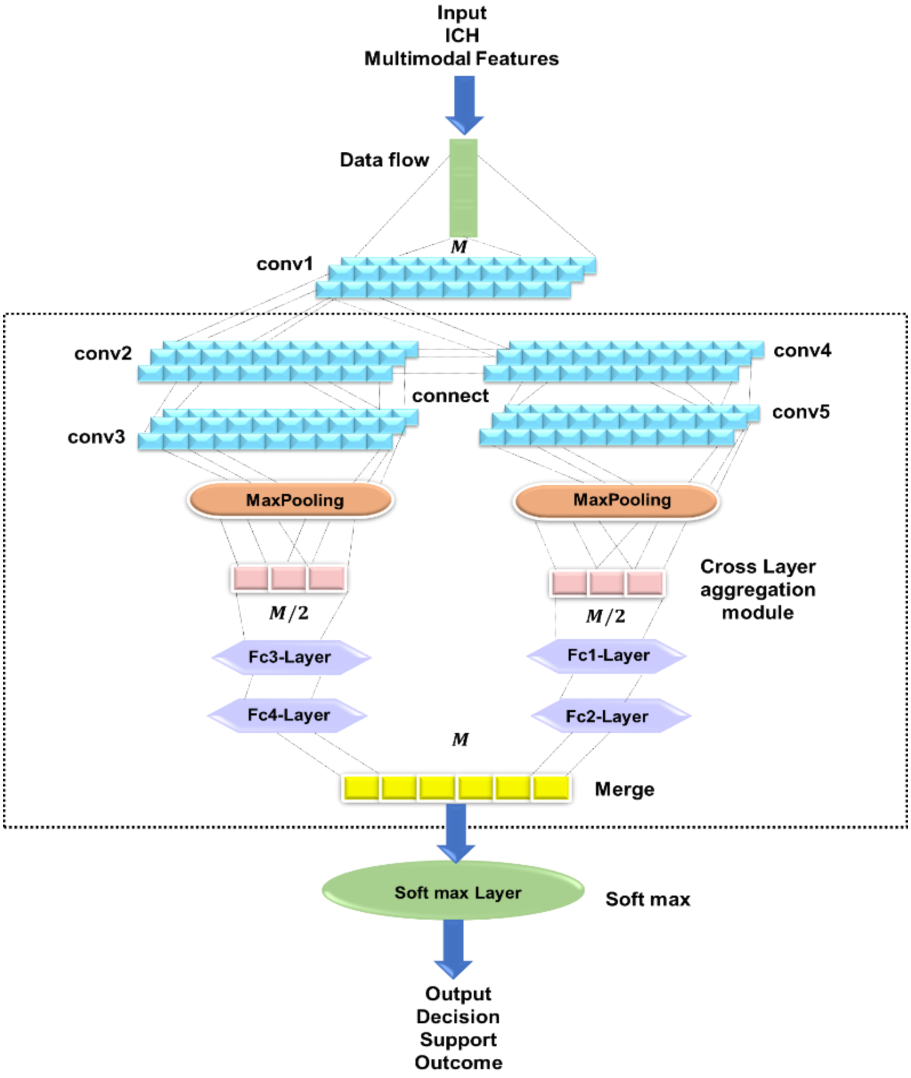


Fig.5: ACNN Architecture for Cultural Feature Learning

The input layer receives visual data in a matrix of dimensions n×o, where n represents the height and o the width of the input ICH data. This input is processed by the convolutional layer, in which learnable filters of size e×e traverse the data to extract relevant local features, including textures, shapes, and symbolic patterns associated with cultural content. Feature maps are generated through convolution, preserving the spatial relationships inherent in the visual information. To introduce nonlinearity and enhance feature discrimination, an activation function is applied, as defined in Equation (8):

$$\text{ReLU}(a) = \max(0, a) \tag{8}$$

Where, a denotes the input activation value of a neuron. $max(0,\ a)$ outputs zero for negative inputs and the input itself for positive values. This reduces dimensionality, improves computational efficiency, and robustness to spatial variations. The output size of the convolutional layer for an input of size $N \times N$, filter size $e \times e$, and padding q is computed using Equation (9):

$$(N + 2q - e + 1)^2 \tag{9}$$

In processing visual data related to ICH, N denotes the size of the input feature map representing cultural images or videos, e defines the convolutional filter size used to capture meaningful cultural patterns, and q represents padding applied to preserve spatial details, ensuring effective digital management and interpretation of heritage elements. Through this procedure, ACNN ensures stable learning, robust generalisation, and consistent classification of diverse ICH representations.

2.5.2 ILSTM for Temporal Analysis of ICH Data

Figure 6 illustrates the ILSTM model, which processes both delayed and non-delayed inputs from multimodal cultural heritage data. This enables the learning of temporal patterns, adaptive feature correlations, and accurate prediction of ICH, supporting effective prioritisation and resource allocation for heritage management.

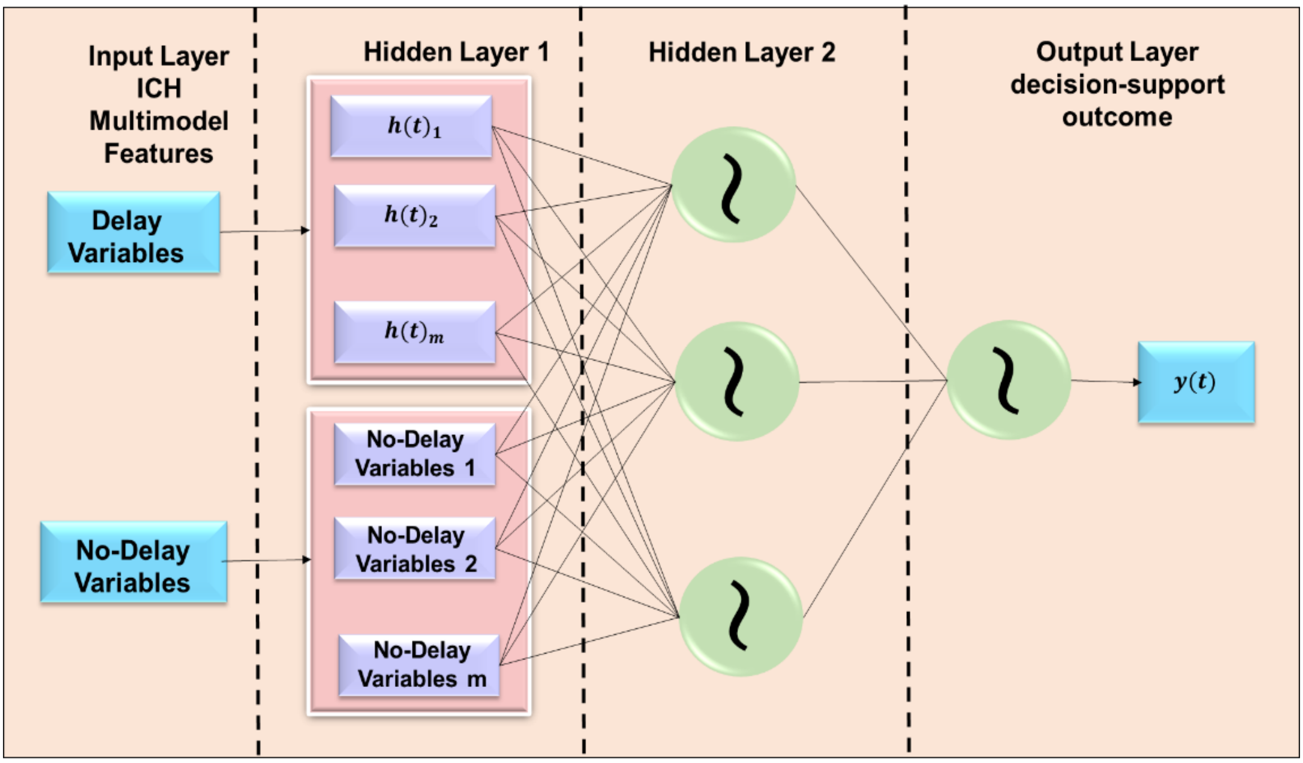


Fig.6: ILSTM Architecture for Multimodal Heritage Prediction

When analysing ICH data, such as sequential patterns in textual narratives or audio recordings of traditional music, conventional LSTM models can lead to reduced classification performance and increased training time, particularly when handling high-dimensional multimodal ICH features. To address this issue, an ILSTM network is proposed, which preserves sequential data while effectively integrating input features. The retained information proportion, referred to as the kept proportion, is defined in Equation (10).

$$\text{Information\_reserve} = \frac{1}{\text{seq\_length}} \times \text{information} \qquad (10)$$

The sequence length (seq_length) denotes the number of consecutive elements in the input, such as words from an oral narrative or frames from a traditional music recording. The information variable represents the total data available at each time step, incorporating all extracted features from textual, audio, or visual modalities. The information_reserve specifies the proportion of this input that is retained and effectively utilised by the ILSTM to ensure accurate classification and feature representation of ICH elements. Equation (11) integrates the sequential input features into a single consolidated vector.

$$X = [x_1, x_2, \dots, x_{\text{seq\_length}}] \qquad (11)$$

Where X denotes the vector at the time step, including textual, audio, or visual embeddings.

The output of the ILSTM network is defined as in Equation (12).

$$\text{Ilstm\_out} = [h_t] \qquad (12)$$

Here, h_t is the final ILSTM cell's hidden state, now enriched by all sequential inputs without discarding intermediate information.

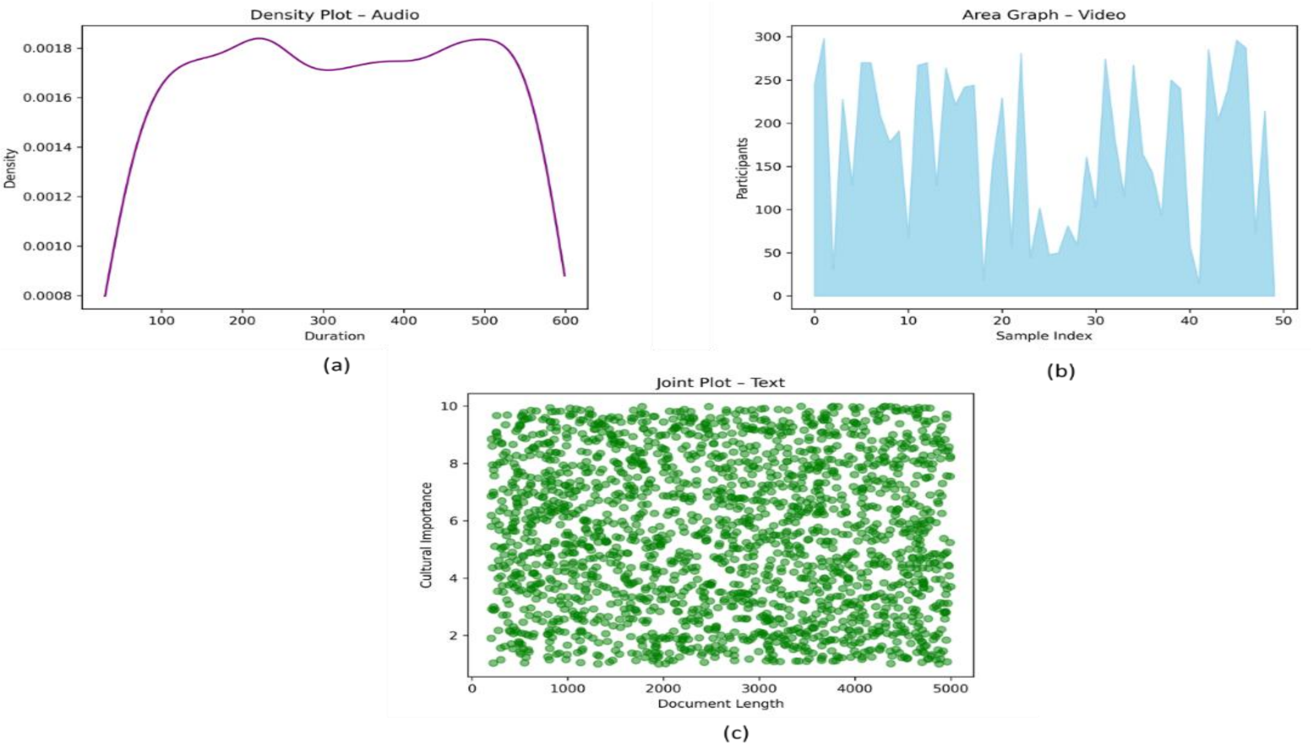
This enhanced design leverages the sequential nature of ICH features to minimise interference, preserve semantic and temporal patterns, and accelerate training. It enables effective grouping and prioritisation of ICH elements by identifying dynamic trends within the data. The hybrid framework integrates ACNN and ILSTM within ACSIMN, utilising the multimodal properties of visual, audio, and textual ICH information to capture spatial structures alongside temporal-semantic relationships while maintaining cultural richness. This approach supports informed decision-making for sustainable ICH management, improving conservation, accessibility, and long-term engagement with cultural assets.

3 Result

This section presents the results of comparing the ACSIMN-based framework with existing intelligent and multimodal frameworks, demonstrating that the proposed framework outperforms others in recognising and prioritising ICH assets. Assessment and classification of ICH were conducted using the ACSIMN framework implemented in Python 3.11 on an Intel Core i7-12700H, enabling efficient multimodal feature extraction, sequential modelling, and intelligent decision support.

3.1 Performance Validation of Proposed Framework

The predictability, convergence, and stability of the proposed framework are examined using multimodal cultural datasets. The training strategy is optimisation-based, enhancing performance stability and minimising error propagation. Cross-class evaluation demonstrates comparable recognition performance across different types and modalities of ICH. The findings indicate robust generalisation and effective multimodal feature learning, suitable for deployment in mobile and VR-based decision-support applications. Figure 7 illustrates the distribution and characteristics of the multimodal dataset employed for ICH analysis, showing (a) audio duration density patterns, (b) variability in video and image contributions, and (c) the relationship between textual length and cultural significance.



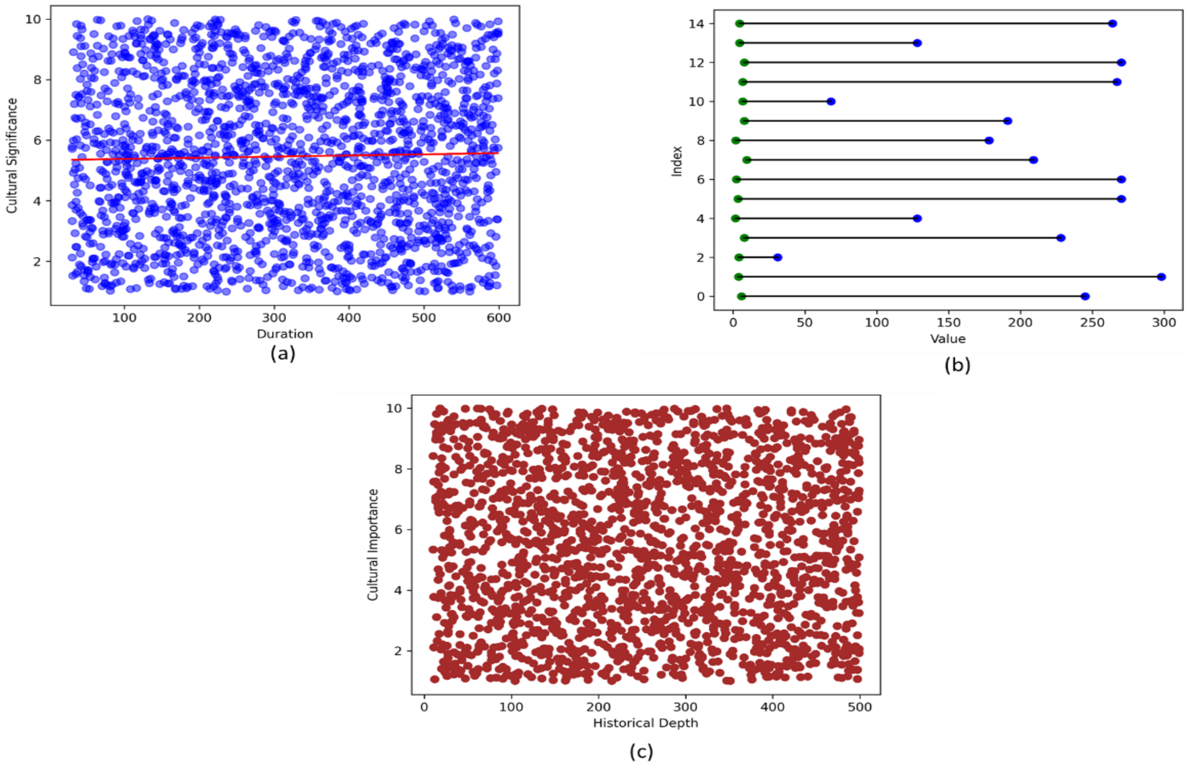
(a)

(b)

(c)

Fig.7: Illustration of Multimodal Data Characteristics (a) Audio Duration Density Distribution, (b) Video and Image Trends, and (c) Textual Document Length Versus Cultural Importance Relationships

Figure 8 depicts trends in multimodal relationships across cultural data modalities, emphasising correlations between (a) audio duration and cultural significance, (b) comparative value patterns in video data, and (c) textual historical depth in relation to cultural importance.



(a)

(b)

(c)

Fig.8: Multimodal Relationship Analysis: (a) Audio Duration Versus Cultural Significance Regression, (b) Video and Image-Based Value Comparison across Samples, and (c) Textual Historical Depth Versus Cultural Importance Distribution

3.2 Performance Evaluation

The following section presents a comparison between the proposed ACSIMN framework and existing intelligent multimodal learning approaches, including CNN combined with Bi-LSTM and Transformer [17] and the Artificial Lizard Search–optimized Multi-Kernel Support Vector Machine with LSTM (ALS-MKSVM-LSTM) [15]. The performance of intelligent ICH prioritisation and recognition is evaluated using relevant performance metrics.

3.3 Metrics Explanation

Accuracy (%): Represents the overall effectiveness of recognising and prioritising ICH within multimodal cultural datasets. Higher accuracy reflects efficient integration of visual, auditory, and textual features in decision-making processes.

Precision (%): Denotes the proportion of correctly identified cultural assets relative to all assets marked as significant. High precision indicates minimal false prioritisation in heritage management decisions.

Recall (%): Measures the system's ability to correctly detect all ICH elements. A high recall value indicates comprehensive coverage of diverse cultural manifestations across communities.

F1-Score (%): Provides a balanced assessment by integrating precision and recall in multimodal classification tasks. An increased F1-score demonstrates consistent and reliable decision-making across ICH categories.

Parameters (M): Indicates the total number of trainable weights within the ACSIMN architecture. Fewer parameters signify model compactness, enhancing portability for mobile- and VR-based applications.

Inference Time (ms): Assesses the time required by the system to generate a decision when presented with multimodal input. Reduced inference time enables real-time interaction and immersive VR visualisation.

Memory Usage (GB): Refers to the memory needed to process multimodal data and produce decisions. Lower memory consumption supports scalable, energy-efficient execution on resource-limited platforms.

Table 2 and Figures 9–11 provide a comparison of key performance metrics. Existing multimodal approaches, such as CNN + Bi-LSTM + Transformer Sun et al. [17] and ALS-MKSVM-LSTM [15], perform competitively on ICH datasets.

Table 2
Performance Comparison using Classification Key Parameters

Methods	**Recall (%)**	**F1-Score (%)**	**Memory Usage (GB)**	**Accuracy (%)**	**Precision (%)**	**Parameters (M)**	**Inference Time (ms)**
CNN + Bi-LSTM + Transformer (Sun et al., 2025)	95.2	94.6	6.4	94.8	94.1	18.6	156
ALS-MKSVM-LSTM (Shang, 2025)	96.0	95.7	3.2	97.0	95.5	12.8	88
ACSIMN (Proposed)	98.1	97.9	2.6	98.4	97.8	9.6	61

However, the proposed ACSIMN model outperforms these methods, achieving higher accuracy (98.4%), precision (97.8%), recall (98.1%), and F1-score (97.9%) while demonstrating substantially lower computational complexity. Additionally, ACSIMN utilises fewer parameters (9.6 M), requires less inference time (61 ms), and consumes less memory (2.6 GB), confirming its efficiency for real-time mobile and VR-based ICH decision-support applications. These results demonstrate that the framework effectively captures fine-grained multimodal cultural features, supports robust cross-modal reasoning, and maintains stable predictive performance across a variety of ICH categories.

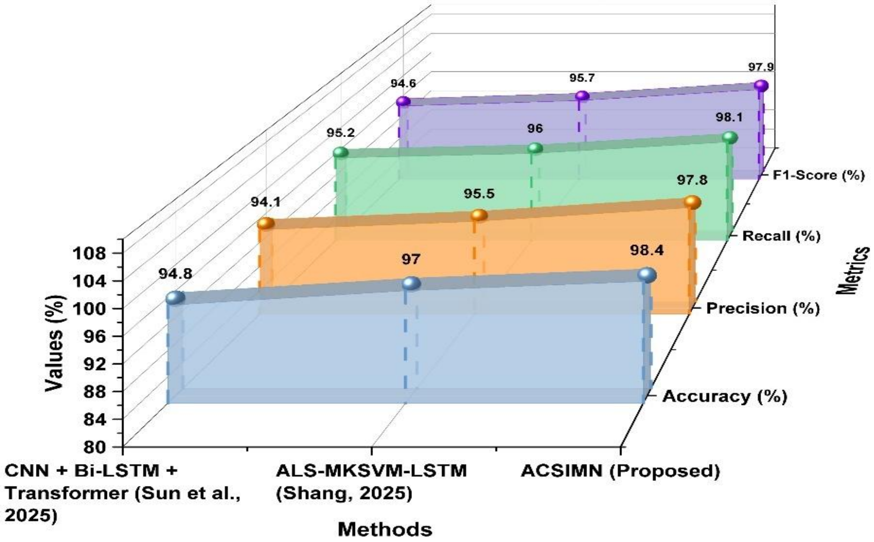


Fig.9: Performance Metric Comparison across Multimodal ICH Models

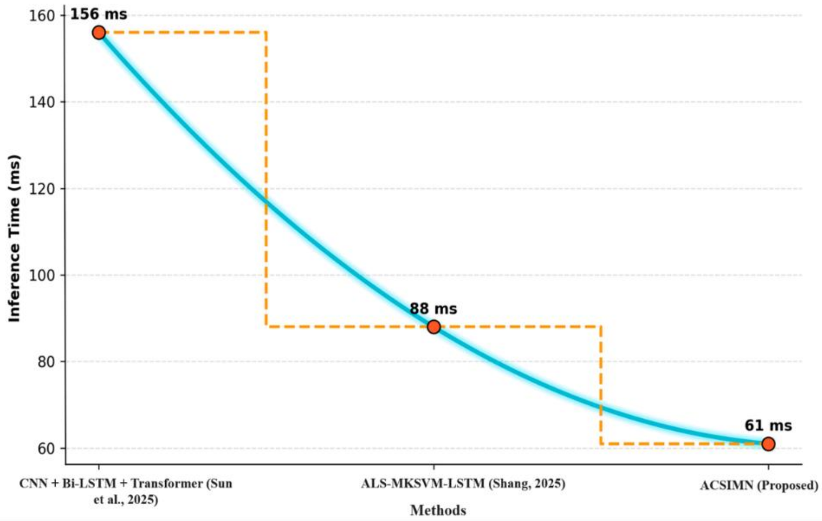


Fig.10: Inference Time Analysis of Intelligent ICH Frameworks

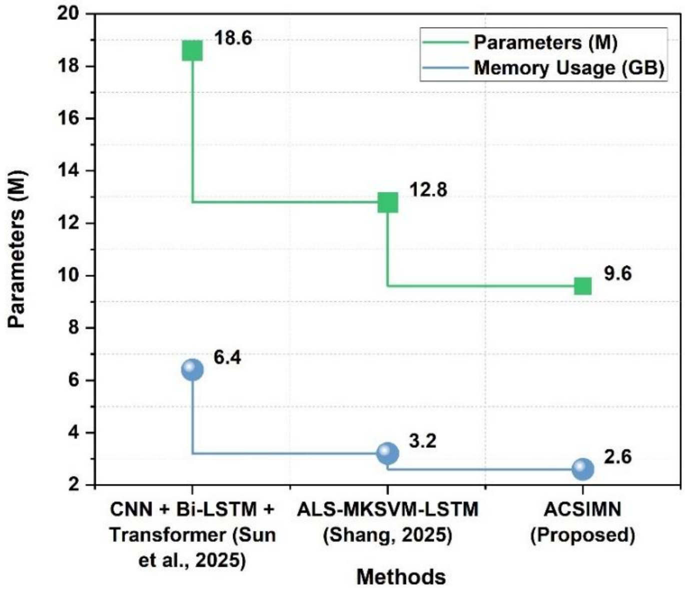


Fig.11: Model Complexity and Memory Usage Comparison

4 Discussion

Exact prioritisation and identification of ICH are challenging due to the diversity of cultural modalities, complex temporal and spatial distributions, and heterogeneous data sources within VR environments. Existing approaches, such as CNN + Bi-LSTM + Transformer Sun et al. [17] and ALS-MKSVM-LSTM [15], achieve moderate success but suffer from high computational demands, limited generalisation, and insufficient handling of fine-grained multimodal features. These limitations can result in partially or wholly inaccurate analyses and prioritization of cultural assets. The proposed ACSIMN framework addresses these challenges by integrating ACNN for spatial feature extraction with ILSTM for temporal and semantic learning, combined through multimodal feature fusion. The framework is further enhanced with VR-based accessibility and immersive visualisation, enabling real-time exploration, evaluation, and management of cultural resources. This integrated methodology ensures accurate identification, effective cross-modal reasoning, and scalable, interactive, real-time decision-support to promote sustainable ICH management.

This study contributes to the discourse on ICH management by introducing a smart decision-support system that merges mobile and VR technologies. The findings align with recent research emphasising the growing impact of digital tools in enhancing participation, governance, and sustainability in heritage management. Specifically, the framework responds to the need for adaptive, data-driven, and stakeholder-inclusive management in complex socio-cultural contexts. In line with the multi-stakeholder approach proposed by Zhang et al. [22], the framework positions digital technologies not merely as exhibition tools but as mediating infrastructures that connect heritage professionals, policymakers, tourists, and local communities. It facilitates interactive engagement and co-creation, strengthening audience interest and cultural relevance by embedding mobile and VR technologies in decision-making. This participatory approach enhances the experiential and interpretive dimensions of ICH, moving beyond static representation toward dynamic cultural interaction.

Operational Perspective: The framework draws on digitalisation workflows identified by Nguyen et al. [14], emphasising systematic acquisition, processing, and management of heritage data. Unlike linear digital documentation models, this approach incorporates real-time mobile inputs and VR simulations in decision-making, enabling heritage managers to explore alternative strategies more efficiently. This capability improves scenario analysis, monitoring, and adaptive planning, which are critical for managing living heritage practices susceptible to social and environmental changes. The experiential design is consistent with the affective and perceptual aspects of ICH tourism; as [16] note, emotions such as awe significantly influence tourist appreciation and valuation of intangible heritage. By leveraging VR-based immersive storytelling and mobile-mediated interpretation, the framework enhances emotional engagement while generating user interaction data for strategic decision-making. This dual function strengthens both visitor experience and managerial insight, bridging a critical gap between experiential design and heritage management.

Moreover, the framework aligns with circularity and sustainability principles emphasised in heritage conservation literature. Tira and Türkoğlu [18] argue that decision-support systems grounded in circular economy logic facilitate integrated and resilient heritage management. By incorporating iterative feedback loops, resource optimisation, and long-term value creation, the framework combines technological innovation with cultural stewardship, which is particularly relevant for ICH where continuity, community participation, and adaptability are more critical than material permanence. The integration of ICT-based systems also resonates with large-scale European heritage initiatives such as INCEPTION and ROCK, which illustrate the transformative potential of digital technologies in enhancing access, understanding, and protection of cultural heritage [5]. While these programmes focus on built heritage, the current research extends their rationale to emphasise immobility, immersion, and intelligent decision-making, demonstrating the flexibility of ICT-driven heritage models across diverse typologies. Furthermore, the framework contributes indirectly to sustainability assessment by facilitating informed decisions that account for environmental and cultural impacts. Foster and Kreinin [7] stress the importance of detailed indicators for heritage interventions under circular economy principles. Although their focus is on built heritage, their findings highlight the necessity of incorporating impact-awareness into management systems. The proposed framework provides a platform for evidence-based assessment, aligning ICH preservation with sustainability objectives.

Overall, this research synthesises digital innovation, experiential design, and decision-support principles into a coherent framework that streamlines ICH management. It addresses gaps in stakeholder engagement, promotes sustainable heritage practices, and offers a scalable model adaptable to various cultural contexts.

5 Conclusion

The primary focus of this research was the development of an intelligent decision-support framework for managing and prioritising ICH. The proposed model, ACSIMN, integrates ACNN for spatial feature learning with ILSTM for temporal and semantic pattern extraction, enhanced through multimodal feature fusion and further strengthened by mobile accessibility and immersive VR visualisation. The framework demonstrates superior performance, achieving recall of 98.1%, F1-score of 97.9%, accuracy of 98.4%, and precision of 97.8%, while exhibiting reduced computational demands, with 9.6 M parameters, 61 ms inference time, and 2.6 GB memory usage, enabling precise real-time identification and prioritisation of cultural assets. Despite its effectiveness, the framework has limitations: it remains computationally intensive and has primarily been evaluated on curated heritage datasets, leaving its generalisability to other heritage contexts uncertain. Future research directions include enhancing computational efficiency, extending applicability to diverse multimedia cultural datasets, integrating explainable AI for interpretability, and facilitating large-scale, real-world deployment within smart cultural heritage management systems.

References

[1] Bi, H., & Nasir, N. B. M. (2024). Innovative Approaches to Preserving Intangible Cultural Heritage through AI-Driven Interactive Experiences. *Academic Journal of Science and Technology*, *12*(2), 81-84. https://doi.org/10.54097/98nre954

[2] Chen, J., & Cao, L. (2024). AI-enabled Rendering Technology for the Construction and Development of Dunhuang Virtual Pavilion. *Frontiers in Computing and Intelligent Systems*, *7*(3), 1-5. https://doi.org/10.54097/yv2hpa48

[3] Chen, W., Zhou, K., Hu, B., Yang, Y., Xu, Y., Zhuoma, D., Zhu, R., Yang, Y., & Pan, J. (2025). Unlocking visitor experiences in cultural heritage sites with SHAP-interpretable AI and social media sentiment analysis. *npj Heritage Science*, *13*(1). https://doi.org/10.1038/s40494-025-02014-0

[4] Cui, L. (2025). Research on the Protection of Intangible Cultural Heritage Based on Virtual 3D Animation Technology. *International Journal of Cognitive Informatics and Natural Intelligence*, *19*(1), 1-23. https://doi.org/10.4018/ijcini.371402

[5] Di Giulio, R., Boeri, A., Longo, D., Gianfrate, V., Boulanger, S. O. M., & Mariotti, C. (2019). ICTs for Accessing, Understanding and Safeguarding Cultural Heritage: The Experience of INCEPTION and ROCK H2020 Projects. *International Journal of Architectural Heritage*, *15*(6), 825-843. https://doi.org/10.1080/15583058.2019.1690075

[6] Di Matteo, E., Roma, P., Zafonte, S., Panniello, U., & Abbate, L. (2021). Development of a Decision Support System Framework for Cultural Heritage Management. *Sustainability*, *13*(13), 7070. https://doi.org/10.3390/su13137070

[7] Foster, G., & Kreinin, H. (2020). A review of environmental impact indicators of cultural heritage buildings: a circular economy perspective. *Environmental Research Letters*, *15*(4), 043003. https://doi.org/10.1088/1748-9326/ab751e

[8] Fu, Y., Shi, K., & Xi, L. (2025). Artificial intelligence and machine learning in the preservation and innovation of intangible cultural heritage: ethical considerations and design frameworks. *Digital Scholarship in the Humanities*, *40*(2), 487-508. https://doi.org/10.1093/llc/fqaf034

[9] Galani, S., & Vosinakis, S. (2024). An augmented reality approach for communicating intangible and architectural heritage through digital characters and scale models. *Personal and Ubiquitous Computing*, *28*(3-4), 471-490. https://doi.org/10.1007/s00779-024-01792-x

[10] He, Q., & Kosenko, D. (2024). Application and innovation of digital exhibition design in the dissemination of Chinese intangible cultural heritage. *Art and Design*(4), 11-21. https://doi.org/10.30857/2617-0272.2024.4.1

[11] Kai, K. (2024). An Exploration of Cultural Programs Contributing to the Digital Conservation of Intangible Cultural Heritage. *Frontiers in Art Research*, *6*(8). https://doi.org/10.25236/far.2024.060803

[12] Lei, Y., Bruno, N., & Roncella, R. (2025). A Data-driven Information Modelling Approach for Cultural Heritage. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLVIII-M-9-2025*, 813-820. https://doi.org/10.5194/isprs-archives-xlviii-m-9-2025-813-2025

[13] Liu, J., Chen, C., Zou, H., Zheng, J., & Gao, D. (2025). Preserving China's Intangible Cultural Heritage through AR/VR storytelling, social media, and AI narratives. *npj Heritage Science*, *13*(1). https://doi.org/10.1038/s40494-025-02265-x

[14] Nguyen, T. A., Do, S. T., Le-Hoai, L., Nguyen, V. T., & Pham, T.-A. (2022). Practical workflow for cultural heritage digitalization and management: a case study in Vietnam. *International Journal of Construction Management*, *23*(13), 2305-2319. https://doi.org/10.1080/15623599.2022.2054268

[15] Shang, C. (2025). Quantitative assessment of the value of intangible cultural heritage art supported by multimodal machine learning. *Discover Artificial Intelligence*, *6*(1). https://doi.org/10.1007/s44163-025-00697-9

[16] Su, X., Li, X., Wang, Y., Zheng, Z., & Huang, Y. (2020). Awe of Intangible Cultural Heritage: The Perspective of ICH Tourists. *Sage Open*, *10*(3). https://doi.org/10.1177/2158244020941467

[17] Sun, J., Khalid, K. A. T., & Kay, C. S. (2025). Deep Learning models for cultural pattern recognition: preserving intangible heritage of Li ethnic subgroups through intelligent documentation systems. *Future Technology*, *4*(3), 119-137. https://doi.org/10.55670/fpll.futech.4.3.12

[18] Tira, Y., & Türkoğlu, H. (2023). Circularity-based decision-making framework for the integrated conservation of built heritage: the case of the Medina of Tunis. *Built Heritage*, *7*(1). https://doi.org/10.1186/s43238-023-00093-1

[19] Wu, S., Liu, J., Zou, H., Yu, T., & Jiang, Z. (2025). Visual Storytelling in Digital Libraries: Design Strategies for Preserving and Accessing Intangible Cultural Heritage. *African Journal of Library, Archives and Information Science*, *35*(2), 85-97. https://doi.org/10.5281/zenodo.17851419

[20] Xie, R. (2021). Intangible Cultural Heritage High-Definition Digital Mobile Display Technology Based on VR Virtual Visualization. *Mobile Information Systems*, *2021*, 1-11.

https://doi.org/10.1155/2021/4034729

[21] Zhang, B., Shi, H., Wang, C., & Gu, M. (2025). Application of immersive VR in the digital preservation of cultural heritage. *Periodico di Mineralogia*, *94*(1).

[22] Zhang, W., Taib, N., & Taib, M. (2025). Reimagining cultural heritage conservation through VR, metaverse, and digital twins: An AI and blockchain-based framework. *PLOS One*, *20*(11), e0335943. https://doi.org/10.1371/journal.pone.0335943

[23] Zhang, Z., & Li, F. e. (2025). Transforming the exhibition experience of intangible cultural heritage in China: a multi-stakeholder approach to service innovation and audience participation. *Asian Journal of Technology Innovation*, 1-39. https://doi.org/10.1080/19761597.2025.2565164

[24] Ziku, M., Teneketzis, A., Alexandridis, G., Christodoulou, Y., Konstantakis, M., Drosopoulos, A., Dritsas, E., Siolas, G., Paximadis, K., & Rousaki, F. (2024). Digital cultural heritage management for local heritage: overcoming barriers to accessibility with regional digital infrastructures. *Journal of Integrated Information Management*, *9*(1), 20-29. https://doi.org/10.26265/jiim.v9i1.38585