Afr. J. Lib, Arch. and Inf. Sc. Vol. 36, No. 1 (January 2026) 1-16
https://doi.org/10.5281/zenodo.20818230

The Role of Libraries in Preserving eSports Heritage: Archival Strategies for Competitive Gaming Records

Yihan Tang[1]*

[1]Faculty of International College, Krirk University, Thailand

**Corresponding Author: Yihan Tang (xiuhan5171@gmail.com)*

Abstract

eSports has developed into a major socio-digital phenomenon, producing extensive streams of gameplay recordings, competitive event logs, player performance statistics, and associated multimedia artefacts. However, the long-term conservation of eSports history is constrained by several challenges, including rapid technological redundancy, intellectual property limitations, and the inherently short lifecycle of professional gaming ecosystems. To address these issues, this study proposes a digital data archival framework that investigates the function of libraries in the preservation and curation of eSports heritage, particularly within competitive gaming records. Within this framework, an eSports Archive Records Dataset comprising 15 video clip files was compiled, with source materials obtained from the Kaggle platform. Prior to analysis, the dataset underwent pre-processing using min–max normalisation to ensure consistency across metadata attributes, file formats, and indexing schemas. Feature engineering was then conducted to capture spatial configurations, temporal sequences, and motion-based patterns, thereby supporting efficient, scalable, and precise archival processes while retaining the interactive and participatory characteristics of eSports environments. For feature optimisation, the Tiki-Taka algorithm (TTA) was implemented to select the most informative attributes, enhancing analytical accuracy and model performance. In addition, deep learning architectures, specifically Convolutional-tuned Bidirectional Long Short-Term Memory networks (Con-BiLSTM), were applied for automated video analysis tasks, including metadata extraction, in-game event recognition, and highlight generation. Experimental evaluation using Python demonstrated strong system performance across multiple indicators: retrieval accuracy (98.78%), accessibility (92.48%), preservation durability (93.57%), storage efficiency (91.72%), precision (96.37%), recall (93.98%), and F1-score (97.66%). Collectively, these results indicate that the proposed digital archival framework substantially enhances the organisation, discoverability, and functional usability of eSports heritage data. Overall, the findings emphasise the critical role of libraries as custodians of digital cultural assets and offer a structured framework for establishing best practices in eSports archival. This approach bridges traditional preservation methodologies, community-driven documentation efforts, and contemporary AI-enabled systems for sustainable digital heritage management.

Keywords: eSports Heritage, Digital Preservation, Library Science, Archival Strategies, Competitive Gaming Records, Deep Learning.

Introduction

Electronic sports, commonly referred to as eSports, has evolved into a major global competitive domain, attracting vast numbers of participants, spectators, and stakeholders worldwide, with millions actively engaged in organised competitive gaming (García-Méndez and de Arriba-Pérez, 2025). The expansion of eSports is further accelerated by advancements in artificial intelligence (AI), optimised communication infrastructures, and cloud computing systems, which collectively support enhanced analytical capabilities for gameplay interpretation, performance assessment, and global audience interaction (Kim et al., 2024).

Within parallel developments, cultural heritage integration in serious games has gained prominence as an effective mechanism for increasing awareness and facilitating heritage promotion, offering interactive platforms that enhance public education and engagement, thereby contributing to the safeguarding and dissemination of cultural assets at a global scale (Zhang et al., 2024). In this context, serious games have become increasingly significant in pedagogical and heritage communication domains, as they provide immersive and participatory learning environments that demonstrate greater effectiveness in fostering awareness and user engagement compared to conventional instructional methods (Esiri, 2024). Furthermore, the preservation of digital and memory-based heritage within digital libraries necessitates innovative preservation strategies that incorporate emerging technologies, robust infrastructure systems, and standardised metadata frameworks to ensure long-term sustainability (Takle, 2026).

In related technological applications, information systems have been developed to support copyright protection in esports-related educational environments, alongside model deployment strategies aimed at enhancing the effectiveness of digital learning systems (Zhao et al., 2025). Additionally, structured esports training frameworks grounded in learning sciences have been proposed to systematise skill development, performance evaluation, and training programme design (Chang et al., 2024). Machine learning-based indicators have also been developed to predict comeback victory scenarios in Multiplayer Online Battle Arena (MOBA) games, utilising reward-based modelling to strengthen analytical and decision-support systems (Lee and Kim, 2025). Complementarily, artificial intelligence applications in MOBA environments have been systematically mapped to identify methodological approaches, datasets, existing challenges, and future research directions within competitive gaming analytics (Costa et al., 2023).

Despite these advancements, the rapid expansion of eSports continues to generate substantial volumes of digital data; however, structured preservation mechanisms remain insufficient. In particular, libraries currently lack comprehensive archival frameworks capable of systematically collecting, managing, and preserving competitive gaming records. This deficiency poses a significant risk of losing valuable cultural heritage, historical gameplay datasets, and scholarly resources essential for documenting the evolution of eSports. Accordingly, the primary objective of this study is to investigate the role of libraries in the preservation of eSports heritage by developing structured archival strategies for the systematic collection, organisation, and long-term accessibility of competitive gaming records, thereby ensuring the continuity and sustainability of digital gaming culture.

Research Contribution

The primary contributions of this study are outlined as follows:

- A digital data archival framework is presented, demonstrating that the eSports Archive Records Dataset integrates min–max normalisation, spatial–temporal–motion feature extraction, and standardised indexing structures to enhance the scalable preservation of eSports records.
- The TTA is applied to optimise feature selection, enabling the identification of the most relevant attributes while improving archival efficiency and overall data analytical performance.
- ConvBi-LSTM networks are utilised to automate metadata extraction, event detection, and highlight generation, thereby supporting precise and intelligent organisation of competitive gaming archives.

The remainder of this paper is structured as follows: Part II covers related work. Part III presents the proposed digital data archival framework, including TTA and ConvBi-LSTM. Part IV discusses the results and findings, while Part V provides the conclusion.

Literature Survey

The Support Vector Machine with Radial Basis Function (SVM-RBF) has been applied to broaden the scientific accessibility of eSports by enabling the publication of both raw and pre-processed datasets derived from StarCraft II competitions. In empirical evaluation, the model achieved an accuracy of 84.8% for single-player prediction and 90.71% for dual-player prediction (Archibald, 2024). However, its applicability remains constrained due to limited generalisation across different eSports genres and restricted interoperability in cross-platform analytical settings (Białecki et al., 2023). In a broader contextual development, an evolutionary economic geography perspective has been used to construct a theoretical model describing the industrialised evolutionary trajectory of China's eSports sector (Horky et al., 2019). While the eSports industry has undergone substantial transformation driven by rapid technological progress (Hamari and Sjöblom, 2017), the emphasis on the Chinese market narrows the external validity of findings, limiting applicability to global markets characterised by diverse regulatory frameworks and economic conditions (Zhu et al., 2024).

In predictive and uncertainty-aware modelling, a Bayesian Regression-enhanced Long Short-Term Memory (BR-LSTM) approach has been introduced for quantifying uncertainty in Quality of Service (QoS) prediction tasks. The model demonstrates competitive predictive accuracy while additionally producing uncertainty estimates that enhance decision reliability in Service-Level Agreement (SLA) monitoring contexts (Jenny et al., 2017). Nevertheless, its performance is influenced by reliance on historical network data and has not been extensively validated across heterogeneous eSports infrastructures, particularly those affected by latency variability (Yang, 2025). In a related application, a Mask Regional-Convolutional Neural Network (Mask R-CNN) has been employed for object detection within real-time strategy gaming environments using human observational data. This automated observer system surpasses both conventional rule-based approaches and manual monitoring techniques; however, its evaluation remains limited to a narrow set of games, thereby constraining adaptability across diverse eSports titles and gameplay contexts (Joo et al., 2023).

Further advances in predictive modelling include a quantitative framework developed for estimating football player market value, where Random Forest (RF) achieved a coefficient of determination (R^2) of 0.95, outperforming prior benchmark approaches (Al-Asadi and Tasdemır, 2022). Despite strong predictive performance, the reliance on simulated gameplay attributes restricts its ability to fully represent real-world performance dynamics and market valuation complexity. In another domain, Multi-Objective Multi-Instance Learning (MOMIL) has been proposed for eSports outcome prediction, achieving up to 95% accuracy in win forecasting tasks (Mora-Cantallops and Sicilia, 2018). However, its computational demands and evaluation on limited datasets reduce scalability and generalisability across broader competitive gaming environments (Birant and Birant, 2023).

In deep learning-based game analytics, a lightweight neural architecture incorporating multi-scale temporal filters (OS-CNN), alongside an open-source Large Language Model (LLM) for automated video game evaluation, has been proposed for real-time event prediction tasks. The system achieves an F1-score of 0.7470 under balanced conditions; however, its effectiveness is dependent on annotated datasets and shows limited validation across multiple game genres (Sheng et al., 2025). Similarly, a feature-based classification approach for determining eSports player skill levels has demonstrated strong performance improvements, achieving 90.1% accuracy (Seabra, 2017). Nonetheless, dataset size limitations and potential feature sensitivity reduce robustness across varied competitive scenarios (Noroozi et al., 2025). In a different exploratory direction, research into physiological and emotional responses in eSports examined facial colour variation among elderly Japanese participants, revealing a correlation between emotional states during gameplay and cheek region colour saturation (Kim et al., 2020). However, the findings are constrained by a small sample size and demographic specificity, limiting broader generalisability (Kikuchi et al., 2025).

In security and privacy-focused systems, a federated learning-based intrusion detection framework has been developed for wireless sensor networks, integrating a Stacked Convolutional Neural Network and Bidirectional Long Short-Term Memory (SCNN-Bi-LSTM). This approach enhances system reliability by enabling collaborative threat detection while maintaining decentralised and privacy-preserving data processing (Bukhari et al., 2024; Jing et al., 2025). In a related direction, a privacy-preserving data publication method based on an information-driven distributed genetic algorithm (ID-DGA) has been introduced. This approach optimises data utility while simultaneously safeguarding sensitive information, enabling secure sharing and management of large-scale distributed datasets across multiple application domains (Ge et al., 2024; Reitman et al., 2020).

Research Gap

Existing literature reveals a notable absence of standardised archival frameworks for the systematic maintenance of eSports datasets and gameplay recordings, with much of the current research primarily oriented towards predictive accuracy, player classification, and automated commentary generation through machine learning and deep learning techniques (Białecki et al., 2023). In addition, the reliance on small-scale, domain-specific datasets and limited cross-game generalisability significantly restricts long-term data accessibility, reuse, and analytical applicability. Consequently, the integration of analytical modelling with robust digital preservation strategies remains insufficient, thereby hindering the sustained documentation and continuity of eSports heritage (Sheng et al., 2025; Yang, 2025). To address this gap, the present study introduces a digital data archival framework grounded in a structured, library-centric archival system designed to support standardised metadata management, long-term preservation, and interoperable access. The framework is operationally aligned with the TTA and ConvBi-LSTM models, ensuring consistent integration between archival processes and advanced analytical components. Furthermore, it incorporates digital repository architectures, preservation protocols, and collection governance policies to enable cross-platform dataset reusability and durable documentation of competitive gaming records, extending beyond isolated machine learning applications and supporting comprehensive eSports heritage preservation across individual game environments.

Methodology

Digital Data Archival Framework

The Digital Data Archival Framework refers to a structured and systematic approach for the collection, organisation, storage, and long-term archiving of game footage, tournament records, and associated metadata, ensuring sustained accessibility and preservation of eSports data. As illustrated in Figure 1, the framework initiates with the acquisition of eSports data from multiple heterogeneous sources, followed by pre-processing and feature extraction stages to standardise and refine the dataset. Subsequently, the Con-BiLSTM model is applied to analyse the most relevant features selected through the TTA, enabling effective information extraction, event detection, and behavioural pattern identification within gameplay sequences. Finally, the processed and structured outputs are stored within a digital library system, which supports efficient retrieval mechanisms and ensures long-term preservation, thereby enhancing the usability, organisation, and sustainability of eSports archival records.

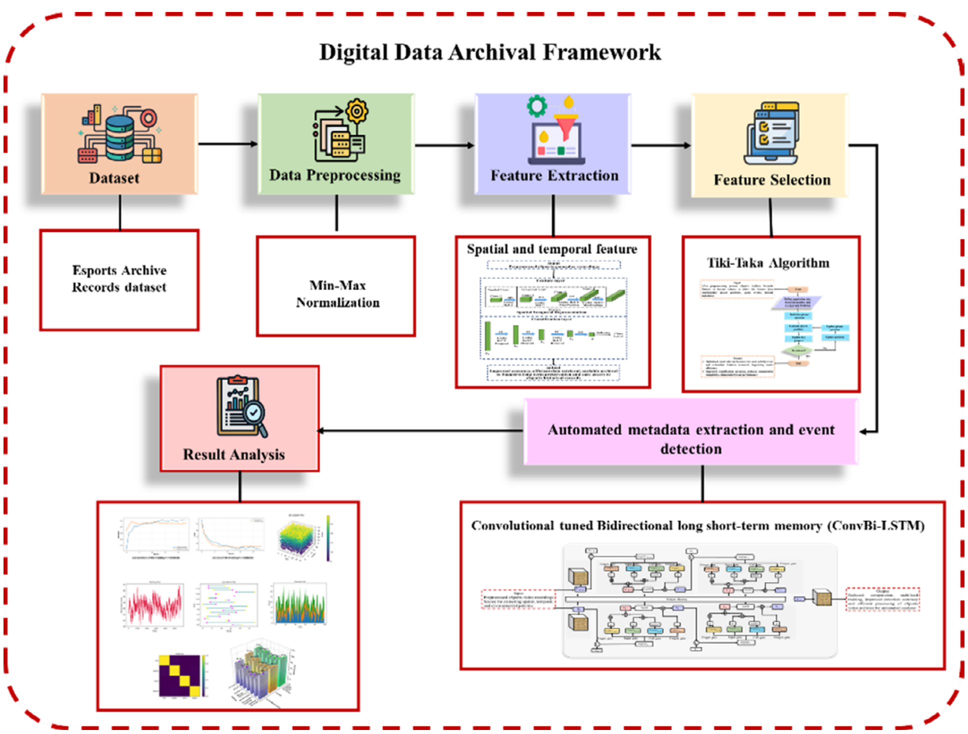


Figure 1: Systematic Representation of the Digital Data Archival Framework.

Dataset

The Digital Data Archival Framework of the eSports Archive Records Dataset was gathered from the Kaggle platform (https://www.kaggle.com/datasets/colabsss/esports-archive-records-dataset). The dataset of competitive eSports gaming records comprises 14 gameplay video files in MP4 format, systematically structured alongside a CSV file containing 44 distinct attributes. These attributes encompass team information, player details, performance analytics, tournament metadata, and multimedia-related features, ensuring a comprehensive representation of competitive gaming events. Each dataset entry is designed to establish a direct linkage between video content and its corresponding archival metadata, thereby supporting systematic event analysis, structured content organisation, improved discoverability, and enhanced accessibility. In addition, this integrated structure facilitates long-term digital preservation of competitive gaming archives by maintaining consistent relational mapping between multimedia recordings and descriptive records. The detailed characteristics of the eSports Archive Records Dataset are presented in Table 1.

Table 1: Features of eSports Archive Records Dataset.

Feature	Description
Record_ID	Unique identifier for each archival entry
Game_Title	Name of the eSports game (e.g., MOBA, FPS)
Tournament_Name	Official name of the tournament
Event_Date	Date when the match or event occurred
Organizer	Organization hosting the tournament
Team_or_Player	Participating team or player name
Match_Stage	Stage of competition (Qualifier, Final, etc.)
Result	Match outcome or winner information
Video_URL	Link to gameplay or broadcast video
Duration	Length of recorded gameplay video
Platform	Streaming platform (YouTube, Twitch, etc.)
Genre	Game genre classification
Region	Geographic location of event
Prize_Pool	Total prize money for tournament
Metadata_Tags	Keywords for indexing and retrieval
Description	Short textual summary of the record
Upload_Date	Date added to archive
File_Format	Video or data format (MP4, CSV, etc.)
Language	Commentary language
Archive_Category	Classification label for archival storage

Data Pre-Processing

Pre-processing was carried out by collecting eSports archival data from multiple heterogeneous sources, including streaming platforms and tournament repositories. During the data cleaning phase, duplicate records, incomplete entries, and corrupted files were removed to ensure dataset integrity, while missing metadata was handled through interpolation techniques and default labelling strategies. To ensure structural consistency, format conversion procedures were applied to standardise file types across the dataset, followed by metadata normalisation to unify naming conventions and attribute representations. Numerical attributes were subsequently rescaled using Min–Max normalisation, enabling uniform value distribution across features. In addition, the indexing structure was reorganised to improve retrieval efficiency and support faster access during analytical operations. Overall, Min–Max normalisation proved particularly effective for eSports datasets due to its ability to preserve relative feature relationships while ensuring consistent scaling across diverse performance variables.

$$W_{norm} = \frac{W - min(w)}{max(w) - min(w)} \quad (1)$$

Here, $max(w)$ denotes the property W's greatest value and $min(w)$ its minimum value. W is an original feature of data value, while W_{norm} is a normalized data value, as shown in equation (2). Following validation, the dataset was confirmed as suitable for downstream processing and subsequently prepared for automated event detection and metadata extraction using the Con-BiLSTM model.

Feature Extraction is used to Extract Spatial and Temporal Patterns in the eSports Archive Records

Feature extraction plays a crucial role in identifying meaningful spatial and temporal patterns within gameplay recordings, thereby enabling accurate organisation, efficient retrieval, and scalable preservation of interactive and collaborative eSports historical data. To minimise feature interference and improve representational stability, the initial spatial convolution layer aligns kernel size with the corresponding channel dimensions. The architecture comprises four sequential feature extraction blocks, each consisting of convolution operations, batch normalisation, leaky ReLU activation, and max-pooling layers to progressively refine feature representations in deeper network stages. Subsequently, temporal convolutions are applied to capture sequential dependencies based on the extracted spatial characteristics, enabling comprehensive spatio-temporal modelling of gameplay dynamics. Finally, fully connected layers integrated with Dropout regularisation are utilised to classify the flattened feature representations. The overall spatial–temporal feature extraction architecture is illustrated in Figure 2.

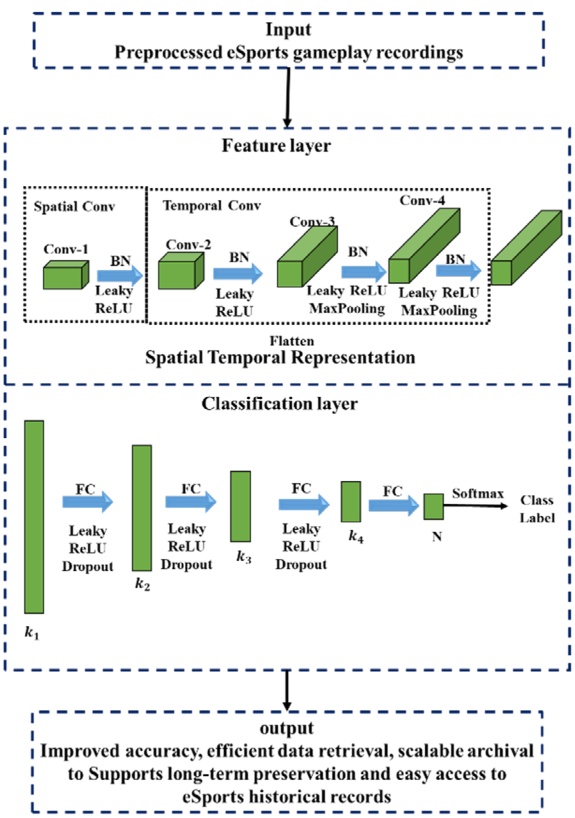


Figure 2: Spatial and Temporal Patterns in eSports Archive Records.

Figure 2 illustrates the Convolutional Neural Network (CNN) architecture, which is structured as a sequential pipeline of convolutional, pooling, and feature extraction layers, followed by dimensionality reduction and classification stages. This hierarchical design enables progressive representation learning, allowing increasingly abstract and informative features to be extracted from archival data for effective analysis and categorisation.

$$l_j = 4l_{j+1,}(j = 1,2,3) \quad (2)$$

Equation (2), l represents the number of neurons or units in one layer of the fully connected classification network. j represents the layer index, which counts layers from the first completely connected layer to the last before SoftMax. The spatial–temporal feature extraction process, which integrates convolutional blocks, regularisation mechanisms, pooling operations, and dropout techniques, enhances model robustness and predictive accuracy. This structured feature learning pipeline supports efficient organisation, rapid retrieval, and scalable archival management of interactive eSports datasets and historical records.

Tiki-Taka Algorithm (TTA) Selects the Most Relevant Features and Removes Redundant Attributes

The TTA was employed for feature selection from the pre-processed eSports archival dataset, where the most significant attributes were identified through iterative evaluation of inter-feature interactions. Inspired by the principle of coordinated passing in team play, the TTA reduces redundant and less informative data while retaining highly discriminative features, thereby improving model efficiency, lowering computational complexity, and enhancing both classification accuracy and retrieval performance. In addition, the framework supports multimodal learning by converting textual match metadata into structured numerical representations. Within this formulation, players and match events are modelled as solution vectors, while short passing interactions and key player contributions are represented as optimisation signals guiding feature selection. This mechanism enables more effective feature prioritisation and overall system optimisation, as illustrated in Figure 3.

Figure 3 represents the Tiki-Taka optimization method initializes players, analyses positions, updates important players using a short-passing approach, iteratively refines solutions, eliminates redundant features, and produces an optimized archival feature subset. O_1, O_2 is the second-dimension value of the first player's position in population matrix. a_1, a_2 is the second-dimension coefficient controlling influence for the first player's update process.

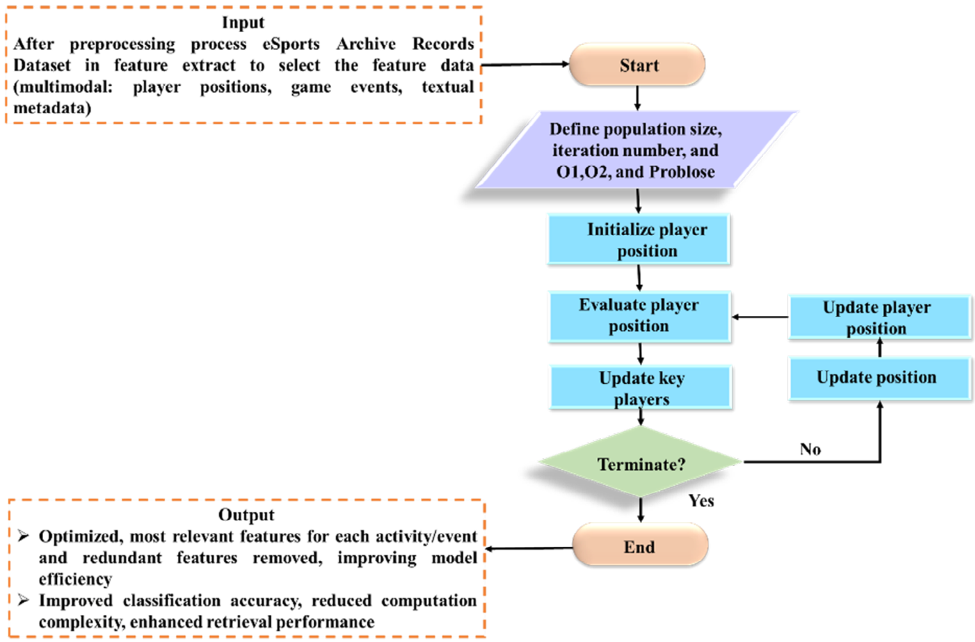


Figure 3: Flowchart of TTA for Optimizing Feature Selection in a Digital Data Archival Framework.

Initially, the player's position is generated at random. The parameters of the algorithm have been defined, including the number of players, task size, critical players, as well as coefficients. In equation (3), a preservation challenge with activities (m) and game events (n). The problem is described using m x n dimensions. O_x and a_x are randomly generated collections of players w with activity dimensions ($m \times n$). This is the possible answer for a population of size x.

$$O_x = \begin{bmatrix} o_{1,1} & o_{1,2} & \dots & o_{1,m\times n} \\ o_{2,1} & o_{2,2} & \vdots & o_{2,m\times n} \\ \dots & \dots & \dots & \dots \\ o_{x,1} & o_{w,2} & & o_{X,m\times n} \end{bmatrix} \quad (3)$$

$$A_x = \begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,m\times n} \\ a_{2,1} & a_{2,2} & \vdots & a_{2,m\times n} \\ \dots & \dots & \dots & \dots \\ a_{x,1} & a_{w,2} & & a_{X,m\times n} \end{bmatrix} \quad (4)$$

In equation (4), A_x represents the movement and velocity of each player in each dimension. Each $a_{X,m\times n}$ change applied to player m in activity/event dimension n. This matrix is utilised to update player positions in the subsequent iteration in accordance with the Tiki-Taka strategy.

An evaluation step is then performed to assess the effectiveness of each candidate solution in preserving and optimising eSports data within the archival framework. Prior to applying the fitness function, player positions are transformed into standardised archival representations to ensure consistency in analytical processing and performance evaluation, as defined in Equation (5).

$$w_j = \begin{bmatrix} w_{1,1,1} & w_{1,1,2} & w_{1,2,1} & w_{1,2,2} & w_{1,2,3} \\ w_{2,1,1} & w_{2,1,2} & w_{2,2,1} & w_{2,2,2} & w_{2,2,3} \\ w_{3,1,1} & w_{3,1,2} & w_{3,2,1} & w_{3,2,2} & w_{3,2,3} \end{bmatrix} \quad (5)$$

If W_j is randomly generated for following iterations; the value is acquired using the TTA method, depending on the new player position from the first iteration. In equations (6) and (7), the matrix element labelled as W_{jmin} is the minimal value from each row in W_j. Then, W_{jmin}. is sorted in ascending order and known as W_j'. Lastly, a position index is used to infer sequence (I). Similarly, at each stage of the archival process, the most suitable feature representation corresponding to each gameplay event is utilised to assign and map the relevant eSports activity to its respective developmental phase, ensuring consistent annotation and structured archival alignment throughout the dataset.

$$W_{j_{min}} = \begin{bmatrix} W_{1_{min}} \\ W_{2_{min}} \\ W_{3_{min}} \end{bmatrix} \quad (6)$$

$$W_j' = ascending\{W_{j_{min}}\} \quad (7)$$

$$I = \{j\};\ \forall i = 1,2,\dots,m \quad (8)$$

In Equation (8), ∀ and m is an assignment applying to all actions (i) in the dataset, and I links each activity to the most relevant event or characteristic for preservation purposes.

$$\widehat{D}_{max} = \frac{D_{max} - D_{max_{min}}}{D_{max_{max}} - D_{max_{min}}} \quad (9)$$

$$\widehat{T}EC = \frac{TEC - TEC_{min}}{TEC_{min} - TEC_{min}} \quad (10)$$

$$\min e = x_1 \widehat{D}_{max} + x_2\, \widehat{T}EC \quad (11)$$

Equations (9) and (10) demonstrate that when D_{max} and TEC have distinct ranges, they must be normalised to the same range [0, 1]. A full test is conducted using one objective function at a time to find the max and min values for each function ($Dmax_{max}$, $Dmax_{min}$, TEC_{max}, TEC_{min}). $\hat{D}_{max}$ and $\hat{T}EC$ represent weightage, respectively. $x1$ and $x2$ denote that both goal functions are equally essential.

One of the TTA features is short passing. The updating method mimics this feature, moving the game from one player to another nearest data point. However, in a genuine game, the opponent may lose possession of the video game while passing. In TTA, the chance of losing the game ($prob_{lose}$). a_j' is the revised object location, and q_o denotes a random integer between 0 and 1.

$$a_{j'} = \begin{cases} rand(a_j - a_{j+1}) + a_j, & q_o > prob_{\text{lose}} \\ a_j - (d_1 + rand)(a_j - a_{j+1}), & q_o \leq prob_{\text{lose}} \end{cases} \quad (12)$$

Equation (12) updates based on two conditions: successful pass ($q_o > prob_{lose}$) and unsuccessful pass ($q_o \leq prob_{lose}$). After a successful pass, simply hand the game to the closest player at randomly. The game is blocked after a failed pass. The coefficient d_1 influences the magnitude of the game reflection.

The player's position is updated in conjunction with the positions of key influential players within the system. Given that the TTA operates with multiple leaders, the selection of important players is performed in a stochastic manner to ensure diversity in guidance signals and prevent deterministic bias. Subsequently, the player's position is updated according to Equation (13).

$$O'_j = O_j + rand * d_2 * (a'_j - O_j) + rand * d_3 * (g - O_j) \quad (13)$$

The Tiki-Taka-based archival architecture, updates each player's position O_j' with both local and global direction. The term (a'_j-O_j) drives the player toward a randomly chosen crucial player (leader), whereas (g-O_j) directs it to the global best solution. O_j is the position vector of the player representing a candidate solution in search space. *rand* is the random number between 0 and 1 controlling stochastic position updates. The effect of these parameters is controlled by random variables and coefficients d_2 and d_3, which balance exploration and exploitation to optimise feature selection and enable efficient, accurate retention of eSports recordings.

Conv-Bi-LSTM Performs Spatial–Temporal Feature for Automated Metadata Extraction and Event Detection

The ConvBi-LSTM model performs feature extraction to capture spatial, temporal, and contextual dynamics from pre-processed eSports video data. In this architecture, the convolutional component is responsible for extracting spatial information such as player movements, game interfaces, scoreboards, and map layouts, while the BiLSTM component models temporal dependencies across frames by analysing sequential relationships in both forward and backward directions to enable accurate event recognition. As illustrated in Figure 4, the ConvBi-LSTM framework integrates convolutional feature extraction with bidirectional sequence learning, incorporating shared weight mechanisms to ensure consistency across temporal representations. This process generates rich contextual embeddings that support effective classification, organisation, and categorisation of archival eSports content.

ConvBi-LSTM integrates bidirectional spatial–temporal learning to enhance representational capacity while reducing high-dimensional computational complexity, mitigating gradient vanishing issues, and improving detection accuracy. Within this architecture, gating mechanisms regulate information flow by suppressing irrelevant inputs and retaining salient features. In addition, convolutional layers replace traditional fully connected weight matrices, enabling more efficient and structured feature learning. The overall model architecture is illustrated in Figure 4.

The forget gate (e_s), the input gate (j_s), and the output gate (p_s) can be obtained by a sigmoid activation function (σ). The cell state (d_s) can be obtained by a hyperbolic tangent activation function (tanh), which is given by equations (14)-(19).

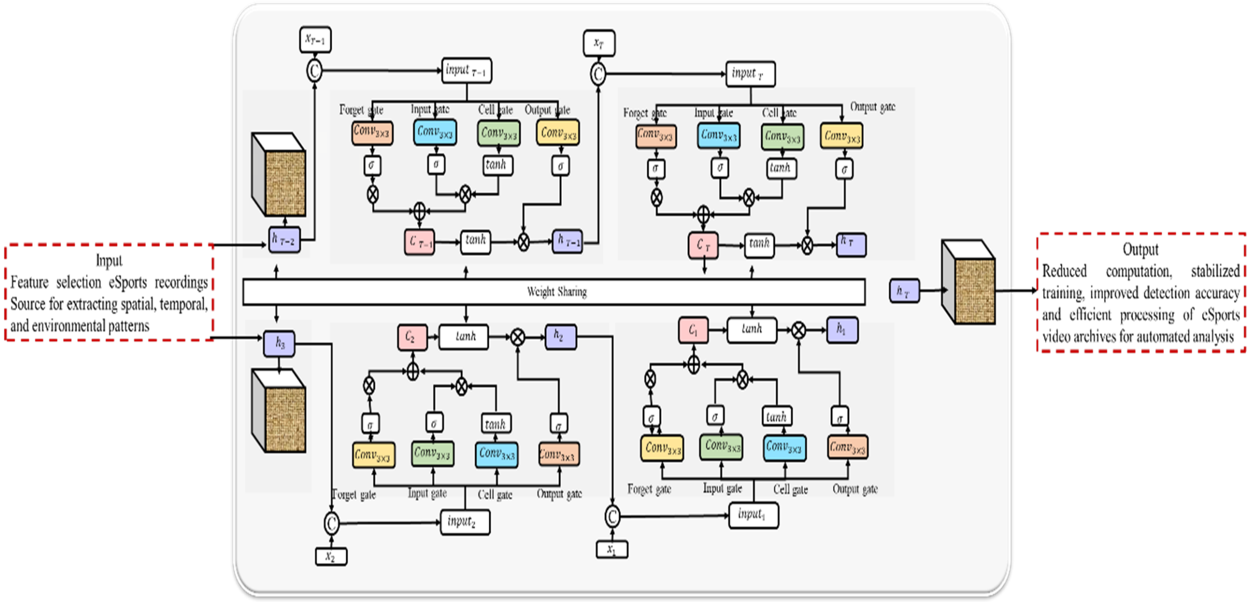


Figure 4: ConvBi-LSTM for Preserving Important Information in Digital Data Archival Records.

$$e_s = \sigma(X_e.[D_{s-1}, g_{s-1}, w_s] + a_e). \quad (14)$$

$$j_s = \sigma(X_j.[D_{s-1}, g_{s-1}, w_s] + a_j). \quad (15)$$

$$d_s = \tanh(X_d.[g_{s-1}, w_s] + a_d) \quad (16)$$

$$D_s = e_s \odot D_{s-1} + j_s \odot D_s. \quad (17)$$

$$p_s = \sigma(X_p.[D_s, g_{s-1}, w_s] + a_p) \quad (18)$$

$$g_s = p_s \otimes \tanh(D_s). \quad (19)$$

The ConvBi-LSTM equations describe gated spatial–temporal feature learning. The forget gate e_s controls how much previous memory D_s-1 is retained, while the input gate j_s regulates new information from the current input w_s. The candidate cell state d_s is covers new features using the tanh activation, and the updated cell state combines retained and new information through element-wise operations. The output gate p_s determines the exposed hidden state g_s, representing bidirectional spatial-temporal features. Convolutional weight matrices X_e, X_j, X_d, X_p, a_d *and* a_p are the bias matrices that are shared across time steps (*s*), enabling efficient learning, while sigmoid and tanh activations control information flow and stabilize gradients. $\odot$ represents element-wise multiplication, and the cell state (d_s) can be obtained by the hyperbolic tangent activation function (*tanh*). $\otimes$ represents element-wise scaling between two tensors. In ConvBi-LSTM, it applies the output gate to the tanh-transformed cell state, producing the final hidden representation. The above two activation functions can be defined as follows, as equations (20) and (21):

$$\sigma(w) = 1/(1 + f^{-w}) \quad (20)$$

$$\tanh(w) = \frac{\sinh(w)}{\cosh(w)} = \frac{f^w - f^{-w}}{f^w + f^{-w}} \quad (21)$$

Here, σ is the Sigmoid activation function; *w* is the input value, and *f* is the exponential base; f^{-w} stands for the exponential decay term controlling output scaling. represents the hyperbolic tangent activation function. *Sinh* (*w*) *and cosh* (*w*) Hyperbolic sine and cosine function. ConvBi-LSTM captures bidirectional spatial-temporal characteristics, filters out unnecessary information via gated activations, stabilizes gradients, and accurately extracts metadata and detects events from pre-processed eSports video archiving data.

The proposed hybrid approach integrates optimisation and deep learning techniques to enhance archival processing efficiency. The TTA is employed to optimise feature selection by eliminating redundant attributes and improving computational efficiency. Subsequently, the ConvBi-LSTM model extracts spatial–temporal information from eSports footage to support automated metadata generation and event recognition. This integrated framework improves classification accuracy, reduces data dimensionality, and strengthens long-term preservation capabilities. Furthermore, the synergy between TTA and ConvBi-LSTM enhances retrieval performance and enables scalable digital archiving of eSports content.

Result Analysis

In the Python-based simulation experiments, performance metrics were utilised to assess the optimal capability and efficiency of feature-driven validation within the proposed system. The effectiveness of the digital data archival framework was evaluated using key performance indicators, including retrieval accuracy, accessibility efficiency, preservation durability, precision, recall, and F1-score.

Experimental Setup

The digital data archival framework for the eSports archival dataset is trained using a standard modelling configuration. The dataset is partitioned into three subsets, comprising 70% for training, 15% for testing, and 15% for validation to ensure balanced evaluation and generalisation. Python 3 is utilised as the primary programming environment for simulation and implementation, while the corresponding hyperparameter settings are detailed in Tables 2 and 3.

Table 2: Hyperparameters of the Digital Data Archival Framework.

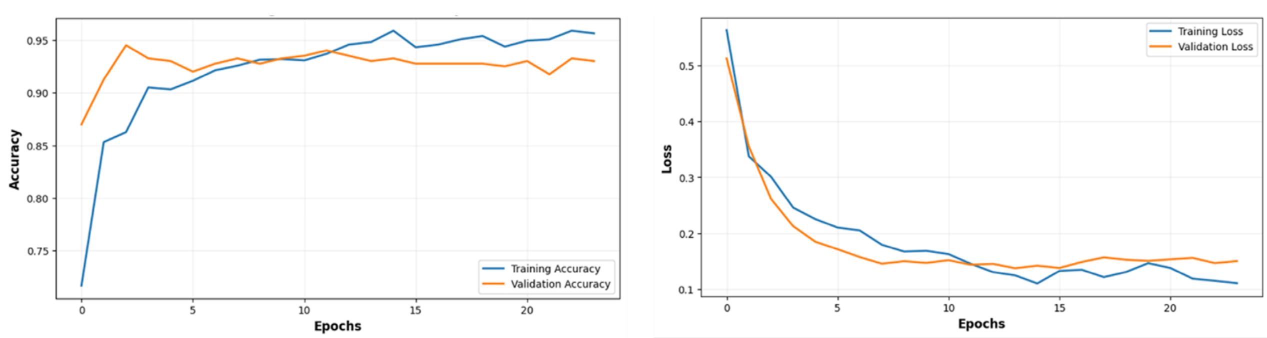
Hyperparameter	Value
Number of Conv Layers	4
Kernel Size (Conv)	3 × 3
LSTM Hidden Units	128
Number of BiLSTM Layers	2
Dropout Rate	0.3
Optimizer	Adam
Learning Rate	0.001
Batch Size	32–64
Epochs	0-20

Table 3: Stimulation Tool of the Digital Data Archival Framework.

Category	Tool / Specification	
Hardware	CPU	Intel Core i7 or AMD Ryzen 7 or higher for efficient computation
	GPU	NVIDIA GTX 1080 / RTX 3060 or higher for accelerated deep learning
	RAM	16–32 GB
	Operating System	Python Windows 10

Performance Evaluation of the Digital Data Archival Framework

The archival classification performance demonstrates progressive improvement during training, indicating reliable effectiveness for digital preservation tasks. As illustrated in Figure 5(a), training accuracy increases from 0.72 at epoch 1 to 0.96 at epoch 23, while validation accuracy remains consistently stable, ranging from 0.92 to 0.94. Similarly, Figure 5(b) shows a steady reduction in loss values, with training loss decreasing from 0.55 to 0.11 and validation loss declining from 0.50 to 0.15. These trends reflect stable model optimisation, improved convergence behaviour, and reduced prediction errors, thereby enhancing the reliability of eSports archival content organisation.



(a) Accuracy of the training vs validation **(b) Loss of the training vs validation**

Figure 5: Digital Data Preservation: Comparing Training and Validation (a) Accurate vs (b) Loss.

The relationship between gamer attributes—specifically kills, assists, and scores—provides a useful basis for generating structured metadata for eSports archival records and indexing within the proposed digital preservation framework. As illustrated in Figure 6, scores range from 0 to 50, assists from 0 to 20, and kills from 0 to 30. The distribution indicates that when kills exceed 20 and assists surpass 10, higher score clusters begin to emerge, reflecting a strong positive association among these performance indicators. The highest density region is observed approximately within 10–25 kills, 5–15 assists, and an average score range of 25–40, highlighting concentrated performance patterns that support more effective archival classification and analytical interpretation.

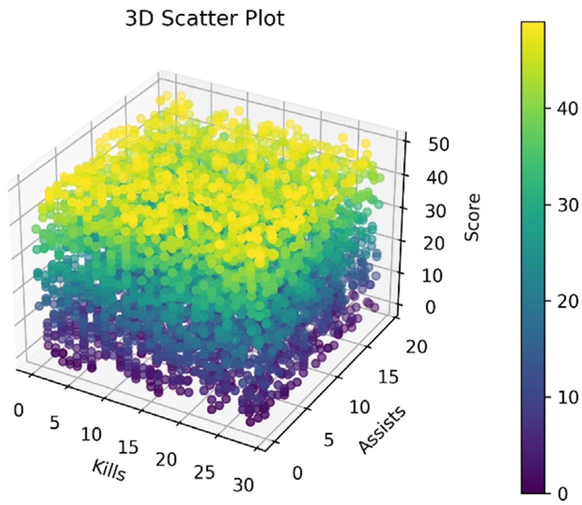


Figure 6: Data Archival Records at a Library are Indexed and Retrieved to Preserve.

Rolling mean trends of gameplay performance are used to support long-term monitoring of archived competitive gaming records and to facilitate the assessment of temporal consistency within eSports datasets. As shown in Figure 7, the rolling mean score fluctuates approximately between 19 and 32 across indices 0–5500. While occasional peaks exceeding 30 are observed around indices 1000, 3000, and 5200, the majority of values remain concentrated within the 23–27 range. Overall, the variance indicates mild fluctuations in performance, while the rolling mean trend remains relatively stable throughout the observed sequence.

The integrated storage and organisation of gaming data, metadata, and audio-visual assets within the digital data archival framework is achieved through the coordinated interaction of multiple archiving components. As illustrated in Figure 8, the three components exhibit individual value ranges from 0 to 30, resulting in aggregated totals reaching approximately 70. When combined values exceed 60, noticeable peaks occur at positions 20, 45, and 65. Overall, the stacked distribution is primarily concentrated within the 30 to 50 range, indicating a relatively balanced contribution from each component and a stable distribution pattern across the archival system.

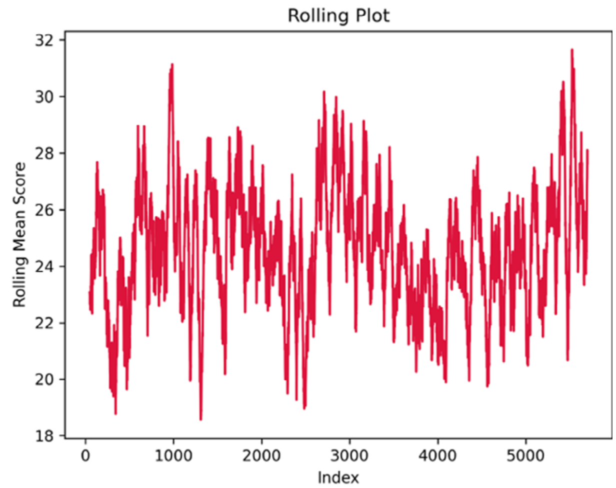


Figure 7: Rolling Mean Score of Gameplay Performance in Digital Data Archival Preserving.

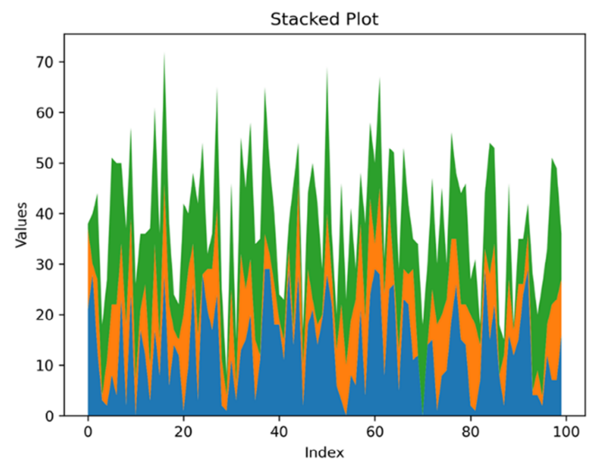


Figure 8: Archiving Components in the Digital Data Archival Framework.

Correlation analysis of gameplay metadata supports relationship modelling, systematic indexing, and structured organisation of competitive gaming records within the digital eSports archival preservation framework. As illustrated in Figure 9, the association heatmap between kills, assists, score, and deaths represents relationship strengths scaled between 0 and 1. These correlation values facilitate detailed metadata analysis by highlighting interdependencies among performance variables, thereby enabling more systematic arrangement, improved interpretability, and enhanced structuring of eSports archival information.

Overall, the performance evaluation indicates that the digital data archival framework achieves robust and reliable indexing, retrieval, and long-term preservation of eSports records. The system demonstrates consistently high accuracy, progressively reduced loss values, stable rolling trend behaviour, well-balanced component contributions, and strong metadata correlations. Collectively, these results confirm the effectiveness and stability of the proposed framework for structured eSports data management.

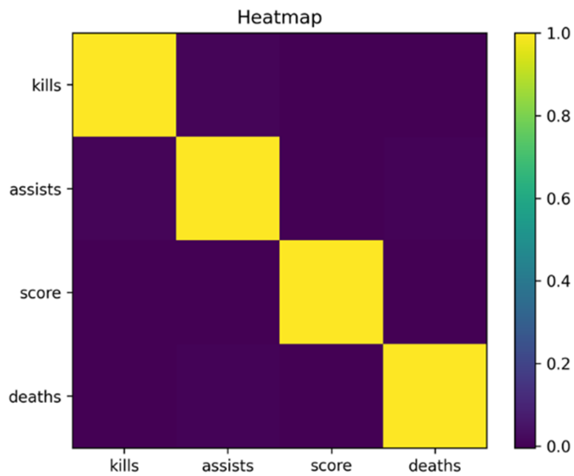


Figure 9: Gameplay Metadata in Digital Data Preservation.

Comparative Analysis

The results indicate that the proposed framework significantly enhances automated metadata extraction and event recognition when compared with baseline approaches. In particular, prior studies such as the SCNN-Bi-LSTM model (Bukhari et al., 2024) and the ID-DGA (Ge et al., 2024) demonstrate federated learning-based digital archival architectures; however, the present framework extends these approaches by integrating structured archival processes with advanced deep learning mechanisms to improve overall archival intelligence and efficiency. Retrieval Accuracy measures the effectiveness of the digital library repository in correctly retrieving eSports records, gameplay videos, and tournament metadata from the archival system. Preservation Durability refers to the system's ability to maintain long-term availability and integrity of competitive gaming records without data degradation or loss. Accessibility reflects the ease with which users can locate and access archived eSports information through structured metadata and systematic indexing mechanisms. Storage Efficiency evaluates the optimal utilisation of digital storage resources while preserving multimedia content, metadata structures, and gameplay recordings.

Furthermore, Precision represents the proportion of correctly retrieved eSports records relative to the total retrieved entries, while Recall measures the system's capability to retrieve all relevant competitive gaming records from the archive. The F1-score provides a balanced evaluation of Precision and Recall, indicating overall reliability in eSports data retrieval performance. Table 4 presents the eSports Archive Records Dataset used for training the standard model and the proposed digital data archival framework.

Table 4: eSports Archive Records Dataset to Train the Standard Model and Digital Data Archival Framework.

Metrics	FL-SCNN-Bi-LSTM	ID-DGA	Digital Data Archival Framework
Retrieval Accuracy (%)	89.88	92.35	98.78
Preservation Durability (%)	80.45	87.21	93.57
Accessibility (%)	79.12	86.54	92.48
Storage Efficiency (%)	76.28	84.63	91.72
Precision (%)	85.23	90.72	96.37
Recall (%)	79.59	86.75	93.98
F1-Score (%)	87.55	94.38	97.66

Table 5: Ablation Study of the Digital Data Archival Framework.

Metrics	CNN only	Conv-Bi-LSTM only	TTA + CNN	TTA + Conv-Bi-LSTM	Digital Data Archival Framework
Retrieval Accuracy (%)	90.42	94.63	95.27	97.31	98.78
Precision (%)	91.18	94.88	95.63	96.82	96.37
Recall (%)	88.67	92.14	92.75	93.46	93.98
F1-Score (%)	89.91	93.50	94.17	95.11	97.66
Storage Efficiency (%)	85.73	88.91	89.84	90.67	91.72
Preservation Durability (%)	86.24	89.76	90.42	92.13	93.57

The Digital Data Archival Framework demonstrates superior performance across all evaluated archival effectiveness metrics. As shown in Figure 10, the system achieves a retrieval accuracy of 98.78%, preservation durability of 93.57%, and accessibility of 92.48%. Storage efficiency is also notably high at 91.72%. In addition, precision, recall, and F1-score improve to 96.37%, 93.98%, and 97.66%, respectively, indicating strong overall reliability in metadata extraction, organisational structure, and discoverability of preserved competitive gaming records. Furthermore, Table 5 presents the ablation study of the proposed digital data archival framework.

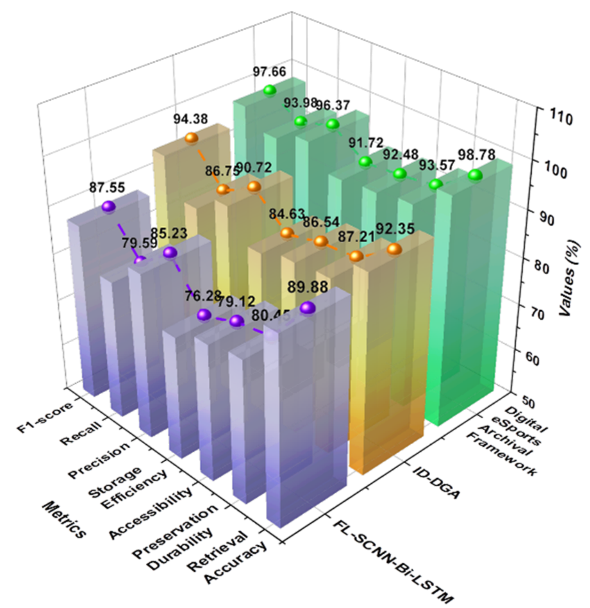


Figure 10: Comparison of the Digital Data Archival Framework and Standard Models.

The ablation study indicates a clear performance improvement from a baseline CNN retrieval accuracy of 90.42% to 98.78% in the complete proposed framework. Correspondingly, precision increases to 96.37%, recall to 93.98%, and the F1-score reaches 97.66%. Additional improvements are observed in storage efficiency (91.72%) and preservation durability (93.57%), confirming the effectiveness of the hybrid integration approach. Overall, the digital data archival system achieves the best performance across all evaluated metrics, demonstrating that the combined framework substantially enhances archival effectiveness, improves metadata extraction accuracy, and strengthens long-term preservation reliability.

Discussion

The digital data archival framework, grounded in a data validation approach, also addresses key preservation challenges identified in prior studies related to eSports data analysis. Existing research relying on specific game databases has shown limited generalisability across diverse eSports genres, alongside high computational complexity in methods such as symbolic transfer entropy, restricted sample diversity, and increased methodological overhead due to nested cross-validation (Noroozi et al., 2025). Similarly, facial colour analysis studies are constrained by narrow demographic coverage focused on older Japanese adults, limited dataset diversity, sensitivity to environmental conditions such as lighting, and the absence of multimodal physiological signals (Kikuchi et al., 2025).

In addition, federated learning-based SCNN-Bi-LSTM frameworks exhibit high communication overhead, dependency on reliable node participation, constrained evaluation conditions, and convergence instability during training processes (Bukhari et al., 2024). The ID-DGA approach also presents limitations, including substantial computational cost, sensitivity to parameter tuning, restricted scalability to large datasets, and incomplete guarantees of privacy protection under adversarial conditions (Ge et al., 2024). To overcome these limitations, the proposed digital data archival architecture integrates the TTA for efficient feature selection with ConvBi-LSTM for spatial–temporal feature extraction. The TTA reduces redundancy and computational overhead by eliminating irrelevant attributes, while ConvBi-LSTM captures complex multimodal temporal patterns from eSports data. This integrated design enables scalable, privacy-aware, and generalisable classification across heterogeneous multi-game datasets. From a practical perspective, the framework offers significant implications for libraries in eSports preservation, including the establishment of standardised metadata structures, AI-driven indexing mechanisms, and secure digital repository systems. Collectively, these capabilities enhance dataset accessibility, support interoperability, and promote institutional collaboration in the long-term preservation of competitive gaming heritage.

Conclusion

This digital data archiving platform contributes to the preservation of eSports heritage and competitive gaming recordings by providing a structured and technology-driven archival solution. The proposed framework integrates min–max normalisation, spatial–temporal feature extraction, the TTA, and ConvBi-LSTM networks for automated metadata extraction and event detection. Experimental evaluation using the eSports Archive Records Dataset demonstrates strong performance, achieving 98.78% accuracy, 92.48% accessibility, 93.57% preservation durability, 91.72% storage efficiency, 96.37% precision, 93.98% recall, and an F1-score of 97.66%. These results indicate that competitive gaming records can be more effectively managed, retrieved, and utilised over the long term through the proposed archival approach. They also highlight the critical role of libraries in supporting digital cultural preservation and establishing robust archival infrastructures for eSports heritage. However, the current implementation is evaluated on a relatively small dataset, and challenges related to metadata consistency remain a limitation of the framework. To address these issues, future research should focus on incorporating larger multi-game datasets, integrating multimodal data sources, developing lightweight model architectures, adopting federated preservation mechanisms, constructing domain-specific ontologies, and strengthening collaboration with library systems. These improvements aim to enhance scalability, interoperability, and international applicability of eSports archival systems.

References

Al-Asadi, M. A. and Tasdemır, S. (2022). Predict the Value of Football Players Using FIFA Video Game Data and Machine Learning Techniques. *IEEE Access*, 10, 22631–22645. https://doi.org/10.1109/ACCESS.2022.3154767

Archibald, S. (2024). Examining the impact of esports program on higher education: A systematic literature review. *Issues in Information Systems*, 25(3): 459–470. https://doi.org/10.48009/3_iis_2024_134

Białecki, A., Jakubowska, N., Dobrowolski, P., Białecki, P., Krupiński, L., Szczap, A. *et al.* (2023). SC2EGSet: StarCraft II Esport Replay and Game-state Dataset. *Scientific Data*, 10(1): 600. https://doi.org/10.1038/s41597-023-02510-7

Birant, K. U. and Birant, D. (2023). Multi-Objective Multi-Instance Learning: A New Approach to Machine Learning for eSports. *Entropy*, 25(1): 28. https://doi.org/10.3390/e25010028

Bukhari, S. M. S., Zafar, M. H., Houran, M. A., Moosavi, S. K. R., Mansoor, M., Muaaz, M. *et al.* (2024). Secure and privacy-preserving intrusion detection in wireless sensor networks: Federated learning with SCNN-Bi-LSTM for enhanced reliability. *Ad Hoc Networks*, 155, 103407. https://doi.org/10.1016/j.adhoc.2024.103407

Chang, Y. K., Chen, K. and Li, Z. (2024). Systematic Framework for Esports Training: Informing eSports Training with the Learning Sciences. *Simulation & Gaming*, 55(1): 82–108. https://doi.org/10.1177/10468781231219938

Costa, L. M., Drachen, A., Souza, F. C. M. and Xexéo, G. (2023). Artificial Intelligence in MOBA Games: A Multivocal Literature Mapping. *IEEE Transactions on Games*, 16(2): 250–269. https://doi.org/10.1109/TG.2023.3282157

Esiri, S. (2024). Cross-Platform Integration Framework for Increasing Accessibility and Engagement in Esports. *International Journal of Multidisciplinary Research and Growth Evaluation*, 5(1): 1318–1326. https://doi.org/10.54660/.IJMRGE.2024.5.1.1318-1326

García-Méndez, S. and de Arriba-Pérez, F. (2025). Explainable e-sports win prediction through Machine Learning classification in streaming. *Entertainment Computing*, 55, 101027. https://doi.org/10.1016/j.entcom.2025.101027

Ge, Y.-F., Wang, H., Cao, J., Zhang, Y. and Jiang, X. (2024). Privacy-preserving data publishing: an information-driven distributed genetic algorithm. *World Wide Web*, 27(1): 1. https://doi.org/10.1007/s11280-024-01241-y

Hamari, J. and Sjöblom, M. (2017). What is eSports and why do people watch it? *Internet Research*, 27(2): 211–232. https://doi.org/10.1108/IntR-04-2016-0085

Horky, T., Baranovskaa, M., Grimmer, C. G., Jakubowska, H. and Stelzner, B. (2019). Television Sport Journalism at the UEFA Euro 2016 Championships: A Comparison of Live Commentary From Four Countries. *International Journal of Sport Communication*, 12(2): 234–259. https://doi.org/10.1123/ijsc.2018-0138

Jenny, S. E., Manning, R. D., Keiper, M. C. and Olrich, T. W. (2017). Virtual(ly) Athletes: Where eSports Fit Within the Definition of “Sport”.

Quest, 69(1): 1–18. https://doi.org/10.1080/00336297.2016.1144517

Jing, S., Awang, M. M., Wan Pa, W. A. M., Pan, P., Memmert, D. and Zhang, Z. (2025). Interpretable machine learning with SHAP for esports performance analysis of professional counter-strike players from 2012 to 2025. *International Journal of Sports Science & Coaching*. https://doi.org/10.1177/17479541251388864

Joo, H.-T., Lee, S.-H., Bae, C.-m. and Kim, K.-J. (2023). Learning to automatically spectate games for Esports using object detection mechanism. *Expert Systems with Applications*, 213, 118979. https://doi.org/10.1016/j.eswa.2022.118979

Kikuchi, R., Shirai, H., Ishizawa, C., Suehiro, K., Takahashi, N., Saito, H. *et al.* (2025). Feature Analysis of Facial Color Information During Emotional Arousal in Japanese Older Adults Playing eSports. *Sensors*, 25(18): 5725. https://doi.org/10.3390/s25185725

Kim, C., Park, J.-H. and Lee, J.-Y. (2024). AI-based betting anomaly detection system to ensure fairness in sports and prevent illegal gambling. *Scientific Reports*, 14(1): 6470. https://doi.org/10.1038/s41598-024-57195-8

Kim, Y. H., Nauright, J. and Suveatwatanakul, C. (2020). The rise of E-Sports and potential for Post-COVID continued growth. *Sport in Society*, 23(11): 1861–1871. https://doi.org/10.1080/17430437.2020.1819695

Lee, J. and Kim, N. (2025). Development of Machine Learning-Based Indicators for Predicting Comeback Victories Using the Bounty Mechanism in MOBA Games. *Electronics*, 14(7): 1445. https://doi.org/10.3390/electronics14071445

Mora-Cantallops, M. and Sicilia, M.-Á. (2018). MOBA games: A literature review. *Entertainment Computing*, 26, 128–138. https://doi.org/10.1016/j.entcom.2018.02.005

Noroozi, A., Hasan, M. S., Ravan, M., Norouzi, E. and Law, Y.-Y. (2025). An efficient machine learning approach for extracting eSports players' distinguishing features and classifying their skill levels using symbolic transfer entropy and consensus nested cross-validation. *International Journal of Data Science and Analytics*, 20(2): 1349–1362. https://doi.org/10.1007/s41060-024-00529-6

Reitman, J. G., Anderson-Coto, M. J., Wu, M., Lee, J. S. and Steinkuehler, C. (2020). Esports Research: A Literature Review. *Games and Culture*, 15(1): 32–50. https://doi.org/10.1177/1555412019840892

Seabra, D. (2017). Portoan Ultra Group Members' Social Representation of Lisbon and Sport Lisboa and Benfica and Its Influence on the Discourses and Practices of the Portoan Ultra Groups and Their Members. *Physical Culture and Sport*, 73(1): 5–14. https://doi.org/10.1515/pcssr-2017-0005

Sheng, X., Yu, A., Zhang, M., An, G., Park, J. and Cho, K. (2025). Integrating Temporal Event Prediction and Large Language Models for Automatic Commentary Generation in Video Games. *Mathematics*, 13(17): 2738. https://doi.org/10.3390/math13172738

Takle, M. (2026). Analogue and digital future imaginaries in an accelerated society: Norway's national library digitalisation programme and the future library artwork. *International Journal of Cultural Policy*, 1–18. https://doi.org/10.1080/10286632.2026.2645036

Yang, C.-F. (2025). Uncertainty-Aware QoS Forecasting with BR-LSTM for Esports Networks. *Information*, 16(12): 1016. https://doi.org/10.3390/info16121016

Zhang, R., Huang, Q., Luo, J., Xu, J. and Pan, Y. (2024). Enhancing User Participation in Cultural Heritage through Serious Games: Combining Perspectives from the Experience Economy and SOR Theory. *Sustainability*, 16(17): 7608. https://doi.org/10.3390/su16177608

Zhao, G., Cheng, Y., He, C., Dong, F., Guo, X. and Gatbonton, R. R. G. (2025). Information Technology Empowers Esports Course Copyright Protection and Teaching Model Innovation Practice. *International Journal of e-Collaboration (IJeC)*, 21(1): 1–21. https://doi.org/10.4018/IJeC.391069

Zhu, Y., Zhao, Y., Yang, Z. and Cheng, Z. (2024). How does external environment impact emerging industrial development: An evolutionary development approach of the e-sports industry in China. *Asia Pacific Viewpoint*, 65(2): 216–230. https://doi.org/10.1111/apv.12401

Yihan Tang is a PhD student in Business Administration at Krirk University, specializing in e-sports and digital marketing. With a strong academic background in marketing, Yihan Tang holds both a Master of Business Administration (MBA) and a Bachelor of Business Administration (BBA) from Krirk University and Bangkokthonburi University, respectively. Since beginning doctoral studies in 2022, Yihan Tang's research has focused on the intersection of digital marketing strategies and the rapidly evolving e-sports industry.