Vol. 3 Issue No. 1 (2025) pp. 1-60

Journal of Intelligent Systems and applied data science (JISADS)

ISSN (2974-9840) Online

We are pleased to publish the 1st issue (Volume 3) of the Journal of Intelligent Systems and Applied Data Science (JISADS). JISADS is a multidisciplinary peer-reviewed journal that aims to publish high-quality research papers on Intelligent Systems and Applied Data Science. Published: **2025-07-14**.

Editor-In-Chief:
Dr. Wasim Ali
Journal of Intelligent Systems and Applied Data Science (JISADS)
Politecnico di Bari, Italy
Editor@jisads.com / editor.jisads@gmail.com

Journal of Intelligent System and Applied Data Science (JISADS)

Journal homepage : https://www.jisads.com

ISSN (2974-9840) Online

ETHICAL AND EMOTIONAL DESIGN CHALLENGES IN HUMAN-DIGITAL TWIN INTERACTION: A SYSTEMATIC REVIEW

Waleed M. A-Nuwaiser[1 *]

*1 * Computer Science Department, College of Computer and Information Sciences, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, 11623, Saudi Arabia., wmalnuwaiser@imamu.edu.sa*

ABSTRACT

This systematic review analyses the ethical and emotional challenges associated with Human-Digital Twin Interaction (HDTI), an emerging area that integrates real-time data modelling and human-centred artificial intelligence. The review synthesizes findings from 39 peer-reviewed studies in healthcare, education, and industrial sectors, highlighting key issues such as data privacy, algorithmic bias, emotional authenticity, and user autonomy. A thematic analysis demonstrates these challenges at the intersection of technical design and human experience, impacting user trust, emotional well-being, and ethical compliance. The review presents a multidimensional framework that connects essential design elements namely personalization, empathy modelling, and explainability with their ethical implications, emotional effects, and practical implementation strategies. This study emphasizes the significance of emotional calibration, participatory design, and ethical auditing as essential mechanisms for ensuring the responsible deployment of HDTI. The review examines not only individual user concerns but also system-level and societal implications, such as institutional trust, social equity, and the cultural formation of emotional norms. The findings highlight the necessity for interdisciplinary collaboration and policy innovation to ensure that HDTI systems are consistent with the principles of transparency, fairness, and emotional integrity. This study seeks to direct subsequent research and influence the development of ethically and emotionally sustainable digital twin technologies.

Keywords: HDTI, ethics, emotional design, explainable AI, participatory design.

1 . INTRODUCTION

1.1. Overview

HDTI represents a significant advancement in human-AI collaboration, characterized by the integration of detailed and dynamically adaptive digital representations of individuals within socio-technical systems. Digital counterparts designed to replicate cognition, emotion, and behaviour are being increasingly utilized in healthcare, industrial, and educational sectors to improve decision-making, personalize services, and support emotional well-being [1], [2]. This advancement raises ethical concerns and emotional complexities that necessitate immediate scholarly and design focus [1], [2].

Recent studies highlight that HDT systems consistently gather and analyse sensitive biometric, behavioural, and emotional data, resulting in risks associated with privacy violations, discriminatory profiling, and algorithmic bias [1], [3], [4]. As Figure 1 illustrates, these challenges manifest differently across domains, with healthcare facing autonomy risks, industry grappling with overreliance, and education confronting cultural insensitivity.

In healthcare, the risk of undermining patient autonomy occurs when AI-driven Human-Digital Twins(HDTs) serve as substitutes for diagnostic or therapeutic decisions [4]. In industrial applications, concerns regarding dependency on automation and the decline of human initiative have emerged, highlighting

issues related to overreliance on AI systems and the reduction of creative agency [2], [5].

Emotional design issues have concurrently gained significant attention. Although HDTs are progressively integrated with affective computing features, their emotional responses frequently exhibit a deficiency in nuance, cultural sensitivity, or authenticity. Studies indicate that users can develop attachments to HDTs through simulated empathy, leading to superficial or misleading emotional experiences [6], [7], [8].

This poses risks of psychological harm, emotional miscommunication, and unclear relational boundaries, especially among vulnerable groups such as patients, children, and the elderly [6], [9], [10].

The study investigates the ethical and emotional problems related to HDTI systems, as well as how human-centred design concepts might be applied to these concerns across other domains. Given the convergence of AI, cognitive science, and ethics, a multidisciplinary approach is required to ensure that HDTI systems are transparent, fair, and emotionally intelligent.

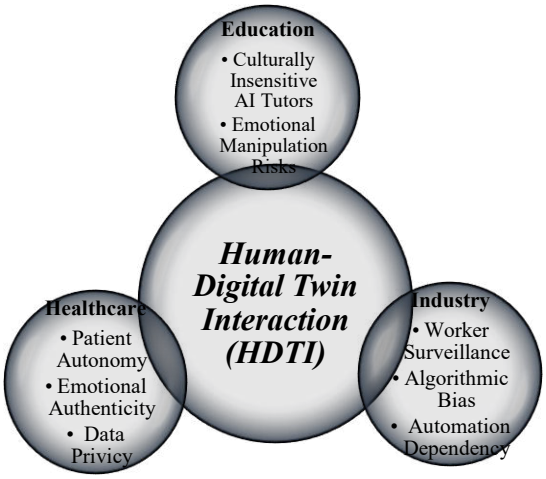

Figure 1: *Key Domains and Ethical-Emotional Challenges in HDTI*

1.2 AI and Human-Digital Twin Interaction

The transition of HDT systems from passive models to interactive, emotionally responsive agents (Emotional authenticity refers to an HDT's capacity to imitate human-like emotional reactions that consumers believe are real [6]) necessitates a focus on emotionally aware and ethically grounded design. Researchers have investigated frameworks for the implementation of trust-building mechanisms, privacy-preserving architecture, explainable AI (XAI), and transparent decision pathways [11], [12]. Affective computing, real-time biometric sensors, and AI-driven behavioural prediction models are examples of technological breakthroughs that enable more adaptable interactions.

Emerging solutions suggest hybrid approaches, such as integrating ethicists in design teams to audit emotional algorithms or adaptive interfaces that adjust transparency levels based on user emotional cues. These innovations highlight the need to balance technical precision with psychological safety in HDTI systems. For example, real-time emotion detection via biometric sensors runs the risk of oversimplifying complex human states (e.g., attributing increased heart rate solely to stress). Similarly, XAI frameworks frequently prioritise technical explainability over emotional intelligibility, leaving users perplexed by "explained" decisions that lack empathetic framing.

1.3 Problem Statement

Research reveals a disconnect between theoretical ethical concepts and their practical validity in dynamic environments such as healthcare and industry. Addressing these gaps is essential for ensuring that HDTI systems do not jeopardise user autonomy or emotional well-being. Despite the existence of normative frameworks for ethical Human-Digital Twin Interaction (HDTI) ([2],[4],[10]), their application does not adequately confront three significant real-world challenges. In healthcare, AI-driven HDTs may prioritise algorithmic "optimisation" over patient preferences, thereby compromising informed consent ([4],[9]). Emotion recognition systems developed on limited datasets often misinterpret cultural and neurodiverse expressions, thereby exacerbating inequalities ([6],[7],[12]). Existing guidelines are unable to adapt to changing contexts (e.g., a patient's declining mental health), thereby increasing the risk of harm ([2],[10]).

In the absence of intervention, these deficiencies are likely to reproduce historical failures of AI characterized by exploitative data practices and emotional manipulation especially among vulnerable populations ([6],[7],[18]). Furthermore, the lack of standardised frameworks for balancing emotional authenticity and user agency is a considerable difficulty.

1.4 Research Question

This systematic review examines the subsequent research question:

1. What ethical and emotional design challenges arise in HDTI, and how can human-centred design principles be applied to address these challenges across various application domains, including healthcare, industry, and education?
2. How can cultural variations in emotional expression and ethical expectations inform the development of more adaptable HDTI frameworks?

2 . Methodology

This systematic review follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [13],[14]. The methodology was designed to capture interdisciplinary perspectives on ethical and emotional dimensions in HDTI systems across healthcare, industrial, and educational domains.

The chosen studies comprise a strategic combination of theoretical, qualitative, empirical, and mixed-methods contributions, enabling the review to utilize both conceptual frameworks and user-centred evidence. The inclusion criteria mandated that each study explicitly addressed ethical considerations, emotional modelling, or both, thereby aligning with the review's objective to investigate emotionally intelligent and ethically responsible HDTI systems. This integrative approach facilitates a thorough analysis of contemporary design strategies, system implementations, and governance challenges within the field. The review process comprises six fundamental stages: (1) Data Sources, (2) Research Strategy, (3) Study Selection Criteria, and (4) Data Extraction Process.

2.1 Data Sources

This review is based on the formulation of the research question: What are the ethical and emotional design challenges in HDTI, and how can these be addressed through human-centred design principles across various application domains, including healthcare, industry, and education? A systematic search strategy was developed and implemented in the Semantic Scholar database, which contains over 126 million academic papers, to thoroughly investigate this question. Additional sources included IEEE Xplore (for technical implementation studies), PubMed (for healthcare-specific applications), and Scopus (for interdisciplinary perspectives).

2.2 Research Strategy

A competent research strategy is essential for refined outcomes following the research questions developed. The research strategy involves the identification and implementation of successful keywords to complete the initial database accumulation of relevant articles.

A total of 499 papers were initially identified, from which 86 studies were selected based on keyword relevance screening. The identified keywords *were "human digital twin," "ethical design," "affective computing," "emotional AI," "human-AI interaction," "healthcare digital twin," and "empathy in AI"* The filtering prioritized studies that specifically examine ethical implications or emotional design within HDTI contexts.

2.3 Study Selection Criteria

Strict inclusion/exclusion criteria were used in the study selection process to ensure methodological coherence. Table 1 below highlights eligibility based on diverse constitutional and study characteristics.

A total of 39 papers fulfilled the criteria. This figure indicates a balance between thematic saturation and the depth of analysis that is manageable, aligning with the size of reviews in related fields such as ethical robotics and affective AI [6], [7].

2.4 Process of Data Extraction

The data extraction phase of this systematic review adhered to a rigorous and methodologically transparent protocol designed to capture the ethical and emotional dimensions of HDTI. This process was developed to address gaps in current HDTI literature, particularly where ethical and affective considerations are frequently neglected, thereby emphasizing the intricate human-centred issues specific to this emerging field. The flow diagram below (Figure 2) indicates the overall mechanism of the final literature selection.

Table 1: *Eligibility criteria for this study selection.*

Criteria	Inclusion	Exclusion
Study Focus	Addresses HDTI as interactive agents rathe r than mere simulations	Mechanical or object twins without HDTI
Design Considerations	Explores ethical or emotional dimensions	Focuses on purely technical implementations
Methodology	Employs qualitative, quantitative, or mixed-methods research	Consists of opinion pieces and editorials
Domain	Pertains to healthcare, industry, or education applications	Involves non-human-centred fields (e.g., robotics)
Language	English only.	Any other language.
Time Frame	2018-2024	Pre-2018 studies

A total of 86 records were initially screened, resulting in a final selection of 39 studies for detailed analysis. This analysis employed a hybrid methodology that integrated AI-assisted semantic categorization with manual thematic validation. A qualitative extraction schema was custom-built, incorporating interdisciplinary concepts from affective computing, human-AI ethics, and responsible design.

A well-structured and relatively systematic review was completed. Scrutiny and the final selection were pertinent to the eligibility considerations set. The illustration (Figure 3) below describes the process of initial research to the final selection stage in the form of a flowchart theme following PRISMA guidelines. Thirty-nine research items were used in the outcomes analysis and quality appraisal at the end.

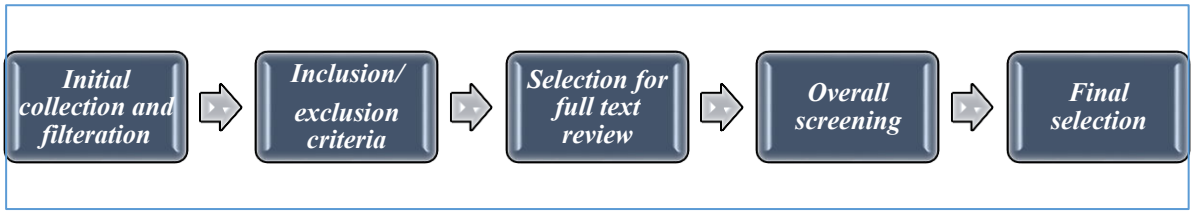

Figure 2: *Summary of the data extraction process (Source: Illustrated by author).*

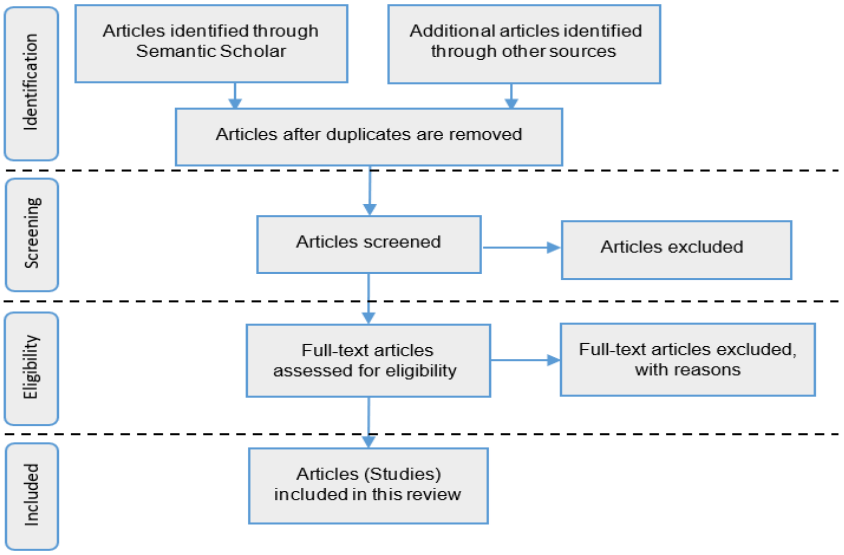

Figure 3: *PRISMA flow chart (Source: Illustrated by author).*

A final dataset of 39 studies was selected for full-text review and thematic analysis, based on methodological rigour and content relevance, from the refined pool. The studies encompass various domains.

The extraction framework included six essential dimensions, detailed below:

- **Study Design and Methodology:** Each study was categorized according to its primary

methodological framework, encompassing qualitative interviews, conceptual frameworks, empirical surveys, experimental designs, or mixed methods. The classifications established a basis for comprehending the depth and breadth of ethical and emotional inquiry in various studies. Qualitative interview studies, such as those conducted by [9], highlighted the perspectives of healthcare practitioners. In contrast, experimental research by [30] examined user responses to HDT-based rehabilitation tools.

- **Application Context:** The specific domains in which each HDTI system was developed or studied were documented, encompassing healthcare, education, industrial, and design environments. The classifications facilitated the identification of domain-specific emotional demands and ethical complexities. Healthcare has become the most prominent field owing to its emotionally sensitive and ethically complex characteristics.
- **Ethical Focus Areas:** Research was categorized based on explicit discussions regarding data privacy, algorithmic fairness, informed consent, transparency, autonomy, and surveillance. [4] explored the ethical implications of surrogate representation of patients, whereas [10] analysed the potential impact of digital twins on human self-understanding and identity, situating these discussions within a wider philosophical examination of ethics, representation, and personhood in healthcare settings.
- **Emotional Integration Strategies**: This category outlines the methods by which systems model, express, or react to emotional signals, including affective feedback loops, adaptive responses, and empathy modelling. [8] developed affect-sensitive interfaces to assess emotional responses, whereas [12] introduced emotionally coherent AI expressions in customer-facing systems.
- **Stakeholder Involvement:** The analysis recorded the extent and manner of involvement of user or stakeholder groups (e.g., patients, designers, caregivers) in co-design processes, pilot testing, or qualitative validation. Participatory studies emphasized the emotional significance and ethical validity contributed by stakeholder input, which was frequently absent in theoretical literature.
- **Design and Governance:** actionable recommendations were documented, including ethical audit protocols, consent management dashboards, and mechanisms for setting boundaries in emotional interactions. The proposals were essential for converting theoretical frameworks into practical safeguards, with research by [2] providing models for participatory governance.

Inter-rater reliability checks were conducted to ensure consistency, with discrepancies resolved through collaborative consensus. Inter-rater reliability assessments were conducted, and discrepancies were addressed through collaborative consensus. This review exclusively incorporates studies that directly address ethical and emotional constructs within the context of HDTI, thereby differentiating it from more general digital twin analyses.

This comprehensive extraction process facilitated a high-resolution synthesis of existing knowledge in HDTI, establishing this review as a key resource for the ethical and emotionally sensitive advancement of future digital twin systems. The extraction process promoted a structured analysis of the emotional and ethical aspects of HDTI by systematically extracting and categorizing relevant elements. The screening phase prioritized the selection of studies based on inclusion criteria, while data extraction allowed for comparative analysis and thematic synthesis, showing broad patterns and practical effects across multiple research disciplines.

3 . Characteristics of Included Studies

This section provides an overview of the 39 studies included in this systematic review. The studies illustrate a diverse and interdisciplinary landscape, highlighting the various domains in which HDTI technologies are being investigated. The analysis indicates a predominant focus on theoretical and conceptual exploration, accompanied by a limited number of empirical and applied studies. The distribution of research types and application contexts highlights the nascent character of the field and the essential requirement for increased applied, real-world studies.

Table 2 presents the characteristics of the included studies, including their methodological design, application domain, and focus on ethical or emotional dimensions. This table offers an overview of the research landscape examined in this review and acts as a reference for the subsequent thematic synthesis.

3.1 Study Types

Classifying studies based on their methodological approaches is essential for a comprehensive

understanding of the research landscape of HDTI. This categorization offers insight into the evolution of the field and the types of evidence that support discussions on emotional and ethical considerations. Study types reflect the diversity of research perspectives, from conceptual debates to empirical validation, and influence the depth of insights regarding user experiences, technological affordances, and ethical risks. The 39 selected studies are distributed by methodological type as follows:

- **Theoretical and Conceptual Studies (30 studies):** These studies offered frameworks, ethical critiques, conceptual models, and analyses of speculative design. Some authors, such as [1] and [4], provided philosophical and normative analyses of digital twin ethics, whereas others, like [2], concentrated on societal risks and emerging disruptions.
- **Qualitative Studies (5 studies):** These investigations utilized interviews, user studies, and interpretive methodologies to analyse perceptions of HDT systems. For instance, [9] analysed the perceptions of medical professionals regarding the incorporation of digital twins into clinical practice.
- **Empirical/Experimental Studies (4 studies):** A limited yet noteworthy collection of studies employed experimental methodologies to assess emotional expressiveness, system responsiveness, or user trust within HDT environments. Significant contributions are found in [6] and [8], which assessed affective interactions and emotional modelling.
- One study employed a **mixed-methods approach**, integrating quantitative and qualitative data to evaluate user responses to emotionally adaptive HDTs within controlled simulations.

3.2 Application Domains

The 39 studies examined encompass various application domains, highlighting the increasing significance of HDTI in multiple contexts. The domains were classified into four primary categories: healthcare, manufacturing/industry, education, and other emerging fields, according to their focus and implementation context.

- **Healthcare (26 studies):** This domain constitutes the predominant portion of the analysed studies. This research focuses on the application of HDTs in personalized care, medical diagnostics, mental health support, and the development of emotionally adaptive virtual agents. Significant ethical issues encompass patient privacy, informed consent, and the genuineness of emotional engagement. Emotional dimensions were particularly significant in therapeutic contexts, where the modelling of empathy and the provision of emotional support were central components. Studies by [7] and [3] illustrate that HDTs can replicate emotional care, while simultaneously highlighting issues related to dependency and the erosion of trust.
- **Manufacturing and Industry (7 studies):** These studies examined the role of HDTs in manufacturing and industry, specifically addressing operator augmentation, productivity optimization, and predictive maintenance. Ethical concerns in this context encompass automation bias, job displacement, and worker surveillance, despite being typically less emotionally intensive. Research, including [5] and [2], examined the impact of HDTs on autonomy and decision-making within smart manufacturing systems.
- **Education (5 studies):** They examined the potential of HDTs to improve learning outcomes via personalized tutoring, emotional feedback, and cognitive engagement. Emotional responsiveness is a critical factor in enhancing student motivation and retention. Ethical considerations encompass data sensitivity, equity in learning analytics, and the potential for emotional manipulation. Examples include AI-driven tutors who can modify their tone and content according to the learner's mood or engagement levels.
- **Other Emerging Fields (2 studies):** A limited number of studies investigated innovative HDTI applications in areas including urban governance, smart mobility, and public policy. The studies primarily concentrated on speculative implementations and conceptual modelling, highlighting issues related to emotional disconnection and social accountability. Although these areas exhibit lower maturity, they indicate potential avenues for the expansion of HDTI beyond conventional sectors.

Table 2: *Characteristics of studies*

Study	Study Type	Application Domain	Human Digital Twin (HDT) Technology Type	Primary Focus
Alimam et al., 2023 [15]	Theoretical/conceptual Analysis	Industry 5.0, Industrial Metaverse	Digital triplet architecture	Integration of Artificial Intelligence (AI) with digital transformation
Arkin et al., 2014 [16]	Theoretical/conceptual Analysis	Healthcare	Robot co-mediators	Preserving dignity in patient caregiver relationships
Bomström et al., 2022 [17]	Qualitative Study	Manufacturing	Human Digital Twins	Design objectives for HDTs in complex systems
Braun, 2021[18]	Theoretical/conceptual analysis	Healthcare	Digital twins in medicine	Ethical implications of digital twins
Braun, 2021 [10]	Theoretical/conceptual analysis	Healthcare	Digital twins in medicine	Ethical challenges of digital twins
Bruynseels et al., 2018[1]	Theoretical/conceptual analysis	Healthcare	Digital twins in personalized medicine	Ethical implications of digital twins
Campanile et al., 2023 [8]	Mixed methods	Healthcare	Emotional aware Human-Machine Interfaces (HMIs)	Inferring emotional models from human machine interactions
Cardin and Trentesaux, 2022 [5]	Theoretical/conceptual analysis	Industrial/ Manufacturing	Human operator digital twins	Ethical implications of HDTs in industrial systems
De Oliveira et al., 2023 [19]	Empirical study (experimental)	Healthcare, Industry	Data-driven emotion modelling for HDTs	Feasibility of emotion modelling for HDTs
El Warraqi et al., 2024[20]	Theoretical/conceptual analysis	Manufacturing	Digital Twin modelling	Human-centricity in manufacturing
Fontes et al., 2024 [2]	Theoretical/conceptual Analysis	Healthcare, Industry/ Manufacturing, Education, Urban Planning/ Governance, X-commerce, Military	Human Digital Twins	Ethical implications and disruptions of HDTs
Gabrielli et al.,	Theoretical/conceptual analysis	Healthcare	Digital twins in digital	Design of AI-powered

2023 [21]			therapeutics	mental health interventions
Garner et al., 2016 [22]	Qualitative study	Healthcare	Virtual carers	Ethical responsibilities in virtual care for the elderly
Hu et al., 2022 [23]	Theoretical/conceptual analysis	Transportation	Driver Digital Twin	Design and enabling technologies for DDTs
Huang et al., 2022 [3]	Theoretical/conceptual analysis	Healthcare	Digital twins for personalized healthcare	Mapping ethical issues of DTs in healthcare
Jabin et al., 2024 [24]	Theoretical/conceptual analysis (scoping review)	Healthcare	Digital health twins	Ethical and quality of care challenges in older care settings
Kabalska and Wagner, 2024 [25]	Theoretical/conceptual analysis	Healthcare, Education, Office work	Human digital Twins	Emergence and impacts of HDTs
Langayan, 2024 [11]	Theoretical/conceptual analysis	Education, Healthcare, Entertainment, Customer service	Digital entities	Establishing genuine human connections through digital entities
Langås et al., 2023 [26]	Theoretical/conceptual analysis	Manufacturing/Industry 5.0	Digital twins for human-robot teaming	Ethical and philosophical implications of DTs in HRT
Lauer-Schmaltz et al., "Beat me if I can!" [27]	Empirical study (experimental)	Healthcare	HDT-based opponents in rehabilitation gaming	Use of HDTs as active elements in serious games
Lauer-Schmaltz et al., 2022 [28]	Theoretical/conceptual analysis (Systematic literature review)	Healthcare	Human Digital Twins	Designing HDTs for Behavior changing therapy and rehabilitation
Lauer-Schmaltz et al., 2024a [29]	Theoretical/conceptual analysis	Healthcare, Workplace optimization	Human Digital Twins	Systematic methodology for designing HDTs
Lauer-Schmaltz et al., 2024b [30]	Empirical study (experimental)	Healthcare	HDTs in rehabilitation	Design and implementation of HDT system for stroke rehabilitation
Lee et al., 2022 [7]	Qualitative study (focus groups with thematic analysis)	Social and emotional interaction with AI	Conversational AI	Emotional bonds between humans and conversational AI
Loveys et al., 2022[12]	Theoretical/conceptual analysis	Healthcare, Customer	Digital humans	Exploring empathy with

	(Scoping review)	service, Education		digital humans
Mandischer et al., 2024 [31]	Theoretical/conceptual analysis	Industrial, Healthcare, Urban Planning	Human Digital Twins	Novel paradigm for modelling humans in human-to-anything interaction
Meghdari and Alemi, 2018 [32]	Theoretical/conceptual analysis	Healthcare, Education, Entertainment/ Gaming	Social & Cognitive Robotics	Ethical challenges in social & cognitive robotics
Mittelstadt, 2021[4]	Theoretical/conceptual analysis	Healthcare	Digital twins in Medicine	Near-term ethical challenges of digital twins
Montag and Diefenbach, 2018 [33]	Theoretical/conceptual analysis	Digital societies and Internet of Things	Digital technologies	Psychological and emotional impacts of digital societies
Nguyen et al., 2024 [34]	Empirical study (experimental)	Emergency response and safety-critical operations	Human Digital Twins in human-AI teams	Trust development in human-AI teams using HDTs
Palmer et al., 2023 [35]	Theoretical/conceptual analysis	Manufacturing and Production	Digital Twin Interface	Symbiotic interface for Digital Twin
Popa et al., 2021 [36]	Qualitative study (interview-based)	Healthcare	Digital twins in Healthcare	Socio-ethical benefits and risks of digital twins in healthcare
Song, 2023 [37]	Theoretical/conceptual analysis	Design	Human digital Twins	Development and impact of HDTs on design
Vildjiounaite et al., 2023 [38]	Theoretical/conceptual analysis	Healthcare (Occupational health and mental wellbeing)	Human Digital Twin	Challenges of learning HDT for mental wellbeing
Wang et al., "Human Digital Twin in Industry 5.0" [39]	Theoretical/conceptual analysis	Manufacturing/Industry 5.0	Human Digital Twin	HDT in the context of Industry 5.0
Wendrich and Kruiper, "Keep IT Real"[40]	Theoretical/conceptual analysis	Design and Human-Computer Interaction	HDT(E) design Tool	Real-time interaction and affective computing in design tools
Xames and Topcu, 2023 [41]	Theoretical/conceptual analysis	Healthcare, Transportation	Digital Twins for Human-in-the-loop	Workload management and burnout

			Systems	prevention in healthcare systems
Zalake, 2023 [9]	Qualitative study	Healthcare	Digital Twins of doctors	Doctors' perceptions of using their digital twins in patient care
de Melo et al., 2023 [6]	Theoretical/conceptual analysis	General human-machine interaction	Virtual humans and social robots	Social functions of machine emotional expressions

These application domains demonstrate the customization of HDTI systems to address the ethical and emotional requirements of various environments. The significance of healthcare studies underscores the necessity of incorporating emotional intelligence into critical interactions, while new fields emphasize the importance of proactive ethical design across various societal sectors.

3.3 Characteristics of Technology and Interaction

The HDTI systems analysed in the studies exhibited considerable variation in complexity and modes of interaction. Some utilized biometric sensors and real-time data streams to replicate human behaviour, while others employed virtual agents and avatars integrated with affective computing algorithms. Emotional interactivity varied from fundamental sentiment detection to sophisticated empathy modelling and expressive feedback mechanisms [12], [11].

Numerous studies have examined trust dynamics, transparency mechanisms, and the relational boundaries between users and HDTs. For example, references [10] and [6] emphasized the psychological risks associated with emotionally manipulative or excessively anthropomorphic human-robot interactions.

These 39 studies collectively offer a comprehensive and nuanced perspective on the HDTI research landscape. Their reflection encompasses the conceptual maturity of ethical and emotional design issues, alongside the urgent requirement for empirical grounding, interdisciplinary collaboration, and context-sensitive design methodologies. The results of this analysis provide a basis for the thematic synthesis and design framework discussed in the subsequent sections.

4 . Thematic Analysis

This section provides a thematic synthesis from the 39 studies reviewed, presenting an integrated perspective on the primary ethical and emotional challenges, along with the associated design implications for HDTI systems. The analysis identified three primary thematic clusters: (1) Ethical Challenges, and (2) Emotional Design Issues. Each theme encompasses interconnected subthemes that represent persistent issues, conceptual conflicts, and actionable priorities in HDTI research (refer to Figure 4).

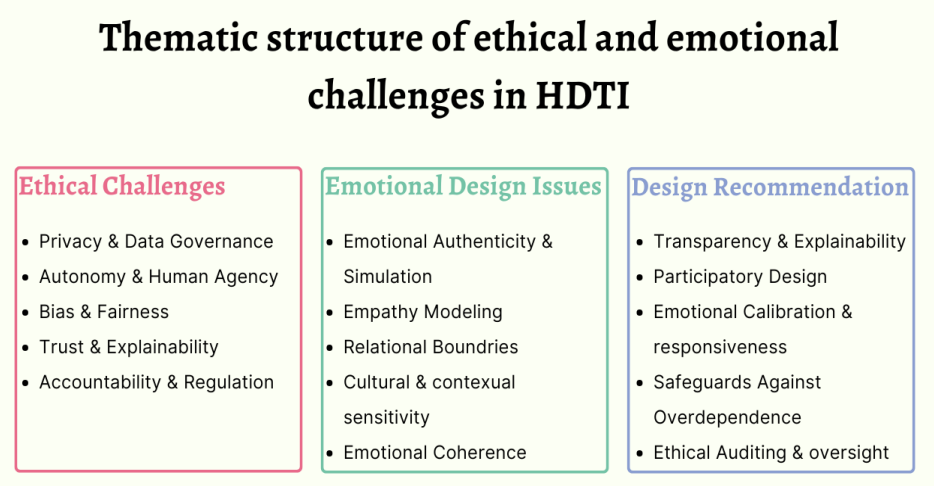

Figure 4: *Thematic structure of ethical and emotional challenges in HDTI*

4.1 Ethical Challenges

Ethical considerations are fundamental to HDTI, particularly given the sensitivity of the data involved and the potential impact of HDTs on users. In the analysed literature, five primary ethical subthemes were identified.

- **Privacy and Data Governance:** numerous studies highlight the risks linked to the extensive collection of biometric, behavioural, and emotional data within the realms of privacy and data governance. Scholars including [1] and [3] emphasize the necessity of robust frameworks for informed consent, data minimization, and access transparency in HDT systems.
- **Autonomy and Human Agency:** the increasing dependence on predictive human decision technologies in decision-making processes, particularly in healthcare and industry, may reduce user autonomy. References [4] and [5] emphasize the ethical considerations associated with the delegation of control to digital replicas.
- **Bias and fairness** are significant ethical concerns, particularly regarding biased data, algorithmic discrimination, and non-inclusive design. Scholars, such as [2], contend that without critical evaluation, digital twins could reinforce existing structural inequalities.
- **Trust and Explainability:** research highlight the significance of fostering user trust via systems that provide clear explanations. References [11] and [12] highlights that transparency in algorithmic processes and emotional responses may reduce user scepticism.
- **Accountability and Regulation**: regulatory oversight and ethical governance are increasingly emphasized in discussions of accountability. [4] and [10] support the need for multidisciplinary collaboration to create policy frameworks that are consistent with ethical HDTI development.

To complement these qualitative themes, Figure 5 quantifies how frequently specific ethical concerns were associated with the ten key HDTI design aspects reviewed in the literature. The graph reveals that privacy risks and bias risks were the most consistently mentioned, each appearing to 2 out of 10 design aspects, reinforcing their prominence across various HDTI applications. Meanwhile, other important concerns such as autonomy loss, trust challenges, deception risk, boundary blurring, empathy challenges, transparency issues, and societal impact each appeared in 1 out of 10 design aspects. This distribution illustrates a fragmented but growing awareness of ethical complexity in HDTI systems. While some issues such as privacy and bias have received sustained scholarly attention, others though equally significant remain underexplored. The graph emphasizes the need for a more integrated and balanced ethical design strategy that ensures these considerations are not treated as isolated risks, but as interdependent elements within a holistic framework of responsible HDTI development.

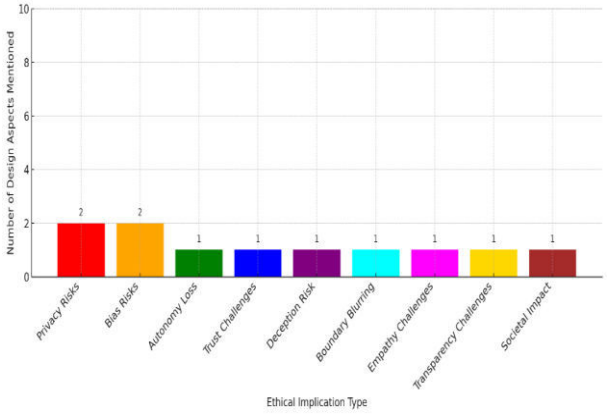

Figure 5: *Ethical design frequency*

4.2 Emotional Design Issues

The emotional dimensions in HDTI systems present distinct design challenges, particularly as these systems strive to recognize, model, and respond to human effects. Five emotional subthemes were identified from the analysis.

- **Emotional Authenticity versus Simulation** presents a persistent challenge, as artificial empathy is often perceived as manipulative or superficial. Cultural norms strongly impact these impressions. Collectivist cultures (e.g., Japan) may allow less expressive HDTs, whereas individualist cultures (e.g., the United States) demand overt emotional participation [12]. Exaggerated expressions might be uncomfortable for neurodiverse users (for example, those on the autistic spectrum), necessitating adaptive calibration [7]. Emotional calibration is the dynamic modification of an HDT's emotional output based on real-time user feedback [8] According to sources [6] and [7], users may encounter difficulties in establishing meaningful connections with HDTs that are devoid of emotional depth or sincerity. [6] examined the advantages and ethical implications of emotional expressions by virtual humans and social robots. [12] underscored the necessity for AI to transcend basic emotion

recognition to attain authentic empathy. [7] observed the potential for superficial human relationships resulting from interactions with AI. Figure 6 illustrates the distribution of emotional implications across ten design aspects identified in the HDTI literature. Among these, negative emotional impacts were the most frequently reported, associated with 4 out of the 10 design aspects, indicating significant concerns around user distress, emotional overload, or discomfort. Mixed emotional impacts, such as simultaneous benefits and risks, and complex emotional impacts, involving nuanced or context-dependent user responses, were each discussed in relation to 2 design aspects. Notably, positive emotional impacts highlighting beneficial emotional outcomes like trust, engagement, or comfort were identified in only 1 design aspect. Similarly, conditional emotional impacts, where user emotions depend on external factors like context or user profile, were also mentioned in just 1 aspect. This distribution underscores a cautious or critical tone in literature, emphasizing the importance of designing HDTI systems that account for a range of emotional outcomes and prioritize emotional safety and coherence.

- **Empathy Modelling:** researchers investigate the role of affective computing in improving empathy simulation. Concerns persist regarding emotional overreach and psychological dependency. These problems are accentuated in cross-cultural settings. For example, HDTs educated on Western data may misinterpret East Asian users' quiet reactions as disengagement, creating prejudice [2]. To facilitate emotional communication in healthcare settings, [16] suggested employing artificial moral emotions in robot co-mediators. [12] highlighted how difficult it is to include genuine empathy in HDTs. A data-driven strategy for implementing empathy was put up by [8], which raised ethical questions about how to manage delicate emotional data.

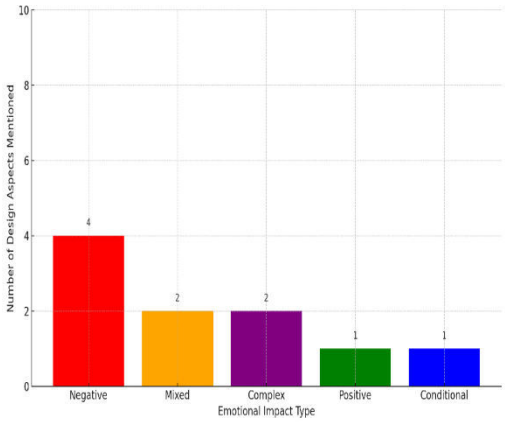

Figure 6: *Emotional design frequency*

- **Relational Boundaries and Dependency:** several studies highlight the importance of maintaining relational boundaries and caution against emotional entanglement between users and HDTs, especially within vulnerable populations. [7] emphasized the capacity of AI to manipulate human emotions and the challenge of differentiating between altruistic and malicious intentions. [9] examined apprehensions about the possible diminishment of doctor-patient contact and the inadequacy of Digital Twins of Doctors for conveying critical information. [18] and [2] examined the notion that HDTs may alter human self-perception and identity.
- **Cultural and contextual sensitivity** is essential, as emotional responses and norms differ among cultures. Emotional reactions differ by culture (e.g., high-context cultures prioritise tone over words [8]), domain (e.g., healthcare requires deeper authenticity than industry [4]), and individual characteristics (e.g., senior users choose clarity over speed [9]). [2] and [3] advocate for adaptive models that consider cultural variations in emotional expression and interpretation.
- **User trust and emotional coherence** are essential for establishing reliability and consistency in emotional responses within HDTs. [12] Emphasize the necessity for emotionally coherent behaviours that align with user expectations.

5 . Design and Implementation Framework

This section presents an integrated design and implementation framework based on the thematic findings, reflecting the key principles identified in the selected literature. This framework, in contrast to purely hierarchical models, captures the multidimensional

relationships among design strategies, their ethical implications, and their emotional impacts within HDTI systems. This analysis derives from the synthesized matrix of findings and highlights the necessity for alignment among values, emotional intelligence, and responsible system behaviour.

The framework is organized as a four-dimensional matrix, integrating insights from the existing literature.

- **Design Aspects** are fundamental system components (e.g., personalization, empathy modelling, explainability)
- **Ethical Implications** are related risks or protective measures (e.g., privacy violations, manipulation, disempowerment)
- **Emotional Impacts** are effects on user experience, encompassing attachment, trust, fatigue, and comfort.
- **Implementation Guidelines** are practical strategies for responsible deployments, such as consent dashboards, co-design sessions, and affective calibration.

The literature presents various human-centred design recommendations to address the identified challenges.

- **Transparency and Explainability:** incorporating explainable AI (XAI) features into HDT interfaces enhances user comprehension of system reasoning, which in turn fosters trust and aligns with ethical standards [11].
- **Participatory Design:** it involves engaging end-users and stakeholders in co-design processes to address ethical and emotional concerns from the outset, as suggested by several studies [2].
- Emotional Calibration and Responsiveness: HDTs must be engineered to adapt emotional expressions dynamically, considering contextual factors, user preferences, and prior interactions [12], [8].
- **Safeguards against Overdependence:** establishing usage limits, relational boundaries, and psychological safety measures is essential to prevent emotional overattachment, particularly in healthcare and educational contexts [9], [7].
- **Ethical Auditing and Oversight:** regular auditing of HDT systems for ethical and emotional impact is essential. This encompasses bias testing, transparency assessments, and inclusive evaluation frameworks [4], [10].

This framework is not static. It is intended as a tool for designers, developers, and researchers to evaluate and adjust HDTI system behaviour at various lifecycle stages. It calls for ethical auditing, participatory co-design, and context-aware emotional calibration as part of implementation planning.

The complete structure of this framework is detailed in the design and implementation matrix provided in **Table 3**. It organizes key design elements alongside their ethical implications, emotional impacts, and recommended implementation strategies.

The detailed matrix provides actionable guidance on balancing innovation with responsibility and aims to ensure HDTI systems are not only technically robust but also socially and emotionally sustainable.

Table 3: *Design and implementation framework*

Design Aspect	Ethical Implications	Emotional Impact	Implementation Guidelines
Privacy and Data Protection	Risk of privacy infringement, data misuse [1]	Potential anxiety and distrust in users [7]	Implement robust data governance, anonymization techniques, and user control over data [24]
Autonomy and Control	Potential loss of human agency, over-reliance on AI decisions [5]	Feelings of disempowerment or loss of self-efficacy [7]	Design for shared control, transparent decision-making processes, and user override options [18]
Trust and Reliability	Challenges in establishing and maintaining user trust [23]	Emotional responses ranging from comfort to scepticism [7]	Ensure system transparency, consistent performance, and clear communication of capabilities and limitations [28]
Emotional Authenticity	Risk of creating shallow or deceptive emotional	Potential for both enhanced emotional	Develop sophisticated emotion models, clearly

	interactions [7]	support and misplaced emotional attachment [7]	communicate the artificial nature of HDT emotions [8]
User-HDT Relationship Boundaries	Blurring of human-machine relationships, potential for over-reliance [7]	Complex emotional responses, potential for confusion or unrealistic [7] expectations	Establish clear guidelines for HDT roles, educate users on the nature and limitations of HDT relationships [11]
Empathy Implementation	Challenges in creating genuine empathetic responses [16]	Enhanced emotional support if successful, risk of perceived insincerity if not [12]	Combine advanced AI techniques with insights from psychology and neuroscience [12]
Personalization	Privacy concerns, potential for bias in personalized interactions [23]	Improved user engagement and emotional connection [27]	Implement adaptive learning algorithms, allow user customization within ethical boundaries [2]
Transparency and Explainability	Difficulty in explaining complex AI decision-making processes [21]	User frustration or mistrust if system actions are not understandable [28]	Develop intuitively interfaces for explaining HDT actions, provide varying levels of detail based on user preferences [28]
Cross-cultural Considerations	Risk of cultural bias or Misunderstanding [3]	Potential for cultural insensitivity or misinterpretation of emotional cues [7]	Incorporate diverse cultural perspectives in design, allow for cultural customization [12]
Long-term Psychological Effects	Potential changes in human self-understanding and social dynamics [10]	Complex long-term emotional impacts on users and society [7]	Conduct longitudinal studies, implement ongoing monitoring and adjustment of HDT systems [33]

6 . System-Level and Social Aspects

Particularly in relation to issues of monitoring, overdiagnosis, automation, social identity, and regulatory uncertainty, the societal effects of HDTI systems are progressively recognized in the literature. These issues are profoundly ingrained in social, institutional, and ethical settings as well as technically based. These ideas are synthesized in the next part with references from current HDTI scholarship.

6.1 Systemic Effects on Institutional Dynamics and Social Trust

The mass acceptance of HDT systems runs the danger of permitting extensive surveillance, therefore compromising democratic principles and institutional confidence, as [2] suggests. While [5] stresses the delicate balance needed between automation and human creativity in industrial settings, [4] raises questions regarding overdiagnosis and the degradation of customized care in healthcare environments. These cases show how accidental HDT technologies while providing efficiency and accuracy, could lower transparency, depersonalize user involvement, and erode relational trust between people and institutions.

6.2 Social Comparisons and Ethical Effects

[25] underline ethical difficulties associated with social inequality by stressing the need for rigorous evaluation of HDTs' ethical impact in different socioeconomic settings. [7] investigated how HDT systems and artificial intelligence might influence generational shifts in relational expectations and value systems. [18] and [2] further highlight how HDTs might drastically change human self-understanding, therefore posing philosophical and ethical concerns concerning identity, authenticity, and digital embodiment. These writers underline together that the implementation of HDTI systems must be context-sensitive to prevent the reinforcement of prejudices and the neglect of weaker groups.

6.3 Social Guidelines and Emotional Conventions

HDTs have complicated emotional and social effects weighted in culture. Emotionally sensitive HDTs could, as [7] propose, cause long-term changes in how society defines empathy, care, and interpersonal interactions. The spread of emotionally expressive artificial intelligence could help to normalize computer simulations of emotion, therefore blurring the distinction between real and synthetic effects. This standardization could minimize emotional variability and change society's view of emotional labour value. Designers must fight the homogeneity of emotional standards and instead support systems reflecting affective plurality.

6.4 Policy, Regulation, And Multi-Stakeholder Engagement Government

The research emphasizes how urgent laws must be developed to control HDT distribution. Particularly in healthcare, [3] underlines the need for organized ethical rules. While [36] supports thorough policy frameworks that manage the sociotechnical complexity of HDT use, [2] warns about gaps in regulation and the absence of accountable procedures. These sources agree that ethical government must be inclusive, initiative-taking, and in line with responsibility as well as innovation and responsibility.

The development of HDTI has enormous potential but also raises major social and systematic issues. HDTI systems must be built and controlled with ethical foresight and a strong dedication to democratic, inclusive ideals if we are to prevent long-term damage and promote favourable results. Collective accountability is essential, as [2] underlines, to make sure that digital twins strengthen rather than undermine the social fabric in which they function. Aligning technology progress with society's well-being depends on regulatory clarity, inclusive design, and continuous review.

7 . Conclusion and Future Directions

This review expands on fundamental issues in HDTI literature, particularly those highlighted by [4] regarding regulatory oversight and [6] concerning risks associated with emotional simulation. This systematic review analysed the ethical and emotional aspects of HDTI within healthcare, industrial, and educational sectors. The review conducted a thematic synthesis of thirty-nine studies, identifying significant challenges concerning data privacy, algorithmic bias, emotional authenticity, and user autonomy. The findings underscore the necessity for design frameworks that are both functionally effective and rooted in ethical considerations and emotional intelligence.

HDTI systems possess significant potential to enhance decision-making, tailor services, and advance human-machine collaboration. As these systems gain autonomy and emotional expressiveness, they introduce novel vulnerabilities, especially in critical areas such as healthcare, eldercare, and mental health support. Emotional simulations devoid of transparency or contextual awareness can result in diminished trust and detrimental dependencies, whereas ethically ambiguous system behaviours pose a threat to individual well-being and public trust.

Future research must focus on creating evaluative frameworks that incorporate ethical auditing alongside affective performance metrics. Empirical studies are essential to investigate user interpretation and responses to the emotional cues of HDTs in various contexts and cultures. Interdisciplinary collaborations among AI researchers, ethicists, designers, and social scientists are crucial for ensuring that HDTI systems embody human values and social complexity. Prior research by [2] and [9] highlights the necessity for inclusive, context-sensitive strategies that are rooted in both domain-specific practices and overarching social norms.

Significant challenges include translating abstract ethical principles into specific design specifications, addressing emotional variance without resorting to stereotypes, and guaranteeing equitable access to technologies that respond to emotional needs. Policies should adapt alongside technological advancements, ensuring a balance between openness to experimentation and the implementation of strong accountability measures.

The success of HDTI systems will rely on technological sophistication as well as their capacity to uphold human dignity, enhance well-being, and facilitate ethical interactions at both individual and societal levels. This review emphasizes the importance of prioritizing emotional intelligence and ethical foresight in the development of future digital twin systems.

References

[1] K. Bruynseels, F. Santoni de Sio, and J. van den Hoven, "Digital Twins in Health Care: Ethical Implications of an Emerging Engi neering Paradigm," *Front. Genet.*, vol. 9, Feb. 2018, doi: 10.3389/fgene.2018.00031.

[2] C. Fontes, D. Carpentras, and S. Mahajan, "Human digital twins unlocking Society 5.0? Approaches, emerging risks and disruptions," *Ethics Inf Technol*, vol. 26, no. 3, Aug. 2024, doi: 10.1007/s10676-024-09787-1.

[3] P. Huang, K. Kim, and M. Schermer, "Ethical Issues of Digital Twins for Personalized Health Care Service: Preliminary Mapping Study," *J Med Internet Res*, vol. 24, no. 1, p. e33081, Jan. 2022, doi: 10.2196/33081.

[4] B. Mittelstadt, "Near-term ethical challenges of digital twins," *J Med Ethics*, vol. 47, no. 6, pp. 405–406, May 2021, doi: 10.1136/medethics-2021-107449.

[5] O. Cardin and D. Trentesaux, "Design and Use of Human Operator Digital Twins in Industrial Cyber-Phy sical Systems: Ethical Implications," *IFAC-PapersOnLine*, vol. 55, no. 2, pp. 360–365, 2022, doi: 10.1016/j.ifacol.2022.04.220.

[6] C. M. de Melo, J. Gratch, S. Marsella, and C. Pelachaud, "Social Functions of Machine Emotional Expressions," *Proc. IEEE*, vol. 111, no. 10, pp. 1382–1397, Oct. 2023, doi: 10.1109/jproc.2023.3261137.

[7] M. Lee, L. Frank, Y. De Kort, and W. IJsselsteijn, "Where is Vincent? Expanding our emotional selves with AI," in *Proceedings of the 4th Conference on Conversational User Interfaces*, ACM, Jul. 2022, pp. 1–11. doi: 10.1145/3543829.3543835.

[8] L. Campanile, R. de Fazio, M. D. Giovanni, S. Marrone, F. Marulli, and L. Verde, "Inferring Emotional Models from Human-Machine Speech Interactions," *Procedia Computer Science*, vol. 225, pp. 1241–1250, 2023, doi: 10.1016/j.procs.2023.10.112.

[9] M. Zalake, "Doctors' perceptions of using their digital twins in patient care," *Sci Rep*, vol. 13, no. 1, Dec. 2023, doi: 10.1038/s41598-023-48747-5.

[10] M. Braun, "Ethics of digital twins: four challenges," *J Med Ethics*, vol. 48, no. 9, pp. 579–580, Aug. 2021, doi: 10.1136/medethics-2021-107675.

[11] R. Langayan, "Establishing Genuine Human Connections Through Digital Entities," *International Journal of Innovative Science and Research Technology (I JISRT)*, pp. 1106–1116, Apr. 2024, doi: 10.38124/ijisrt/ijisrt24apr1280.

[12] K. Loveys, M. Sagar, M. Billinghurst, N. Saffaryazdi, and E. Broadbent, "Exploring Empathy with Digital Humans," in *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstrac ts and Workshops (VRW)*, IEEE, Mar. 2022, pp. 233–237. doi: 10.1109/vrw55335.2022.00055.

[13] A. P. Siddaway, A. M. Wood, and L. V. Hedges, "How to do a systematic review: a best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses," *Annual review of psychology*, vol. 70, no. 1, pp. 747–770, 2019.

[14] A. Liberati *et al.*, "The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration," *Bmj*, vol. 339, 2009.

[15] H. Alimam, G. Mazzuto, N. Tozzi, F. Emanuele Ciarapica, and M. Bevilacqua, "The resurrection of digital triplet: A cognitive pillar of human-machi ne integration at the dawn of industry 5.0," *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 10, p. 101846, Dec. 2023, doi: 10.1016/j.jksuci.2023.101846.

[16] R. C. Arkin, M. Scheutz, and L. Tickle-Degnen, "Preserving dignity in patient caregiver relationships using moral emot ions and robots," in *2014 IEEE International Symposium on Ethics in Science, Technology and Engineering*, IEEE, May 2014, pp. 1–5. doi: 10.1109/ethics.2014.6893414.

[17] H. Bomström *et al.*, "Digital Twins About Humans—Design Objectives From Three Projects," *Journal of Computing and Information Science in Engineering*, vol. 22, no. 5, Apr. 2022, doi: 10.1115/1.4054270.

[18] M. Braun, "Represent me: please! Towards an ethics of digital twins in medicine," *J Med Ethics*, vol. 47, no. 6, pp. 394–400, Mar. 2021, doi: 10.1136/medethics-2020-106134.

[19] C. Dias De Oliveira, A. Khanshan, and P. Van Gorp, "Exploring the Feasibility of Data-Driven Emotion Modeling for Human Di gital Twins," in *Proceedings of the 16th International Conference on PErvasive Technolo gies Related to Assistive Environments*, ACM, Jul. 2023, pp. 568–573. doi: 10.1145/3594806.3596535.

[20] L. E. Warraqi, L. Ragazzini, and E. Negri, "Review on Digital Twin Modelling Applications To Support Human-Centric ity in Manufacturing," in *2024 Annual Modeling and Simulation Conference (ANNSIM)*, IEEE, May 2024, pp. 1–13. doi: 10.23919/annsim61499.2024.10732193.

[21] S. Gabrielli, E. M. Piras, and O. Mayora Ibarra, "Digital Twins in the Future Design of Digital Therapeutics," in *Adjunct Proceedings of the 2023 ACM International Joint Conference on Pervasive and Ubiquitous Computing & the 2023 ACM International Sy mposium on Wearable Computing*, ACM, Oct. 2023, pp. 602–605. doi: 10.1145/3594739.3611325.

[22] T. A. Garner, W. A. Powell, and V. Carr, "Virtual carers for the elderly: A case study review of ethical respons ibilities," *DIGITAL HEALTH*, vol. 2, Jan. 2016, doi: 10.1177/2055207616681173.

[23] Z. Hu, S. Lou, Y. Xing, X. Wang, D. Cao, and C. Lv, "Review and Perspectives on Driver Digital Twin and Its Enabling Techno logies for Intelligent Vehicles," *IEEE Trans. Intell. Veh.*, vol. 7, no. 3, pp. 417–440, Sep. 2022, doi: 10.1109/tiv.2022.3195635.

[24] M. S. R. Jabin, E. V. Yaroson, A. Ilodibe, and T. Eldabi, "Ethical and Quality of Care-Related Challenges of Digital Health Twins in Older Care Settings: Protocol for a Scoping Review," *JMIR Res Protoc*, vol. 13, p. e51153, Feb. 2024, doi: 10.2196/51153.

[25] A. Kabalska and R. Wagner, "Rendering Frankenstein's monsters? The emergence and impacts of human digital twins," *Digital Twins and Applications*, vol. 1, no. 1, pp. 88–92, Sep. 2024, doi: 10.1049/dgt2.12011.

[26] E. F. Langas, M. H. Zafar, and F. Sanfilippo, "Harnessing Digital Twins for Human-Robot Teaming in Industry 5.0: Expl oring the Ethical and Philosophical Implications," in *2023 IEEE Symposium Series on Computational Intelligence (SSCI)*, IEEE, Dec. 2023, pp. 1788–1793. doi: 10.1109/ssci52147.2023.10372069.

[27] M. W. Lauer-Schmaltz, J. Hansen, and K. Kirchner, "'Beat me if I can!' – A Case Study on Human Digital Twin-based Opponen ts in Rehabilitation Gaming," in *Proceedings of the 17th International Conference on PErvasive Technolo gies Related to Assistive Environments*, ACM, Jun. 2024, pp. 277–284. doi: 10.1145/3652037.3652063.

[28] M. W. Lauer-Schmaltz, P. Cash, J. P. Hansen, and A. Maier, "Designing Human Digital Twins for Behaviour-Changing Therapy and Rehab ilitation: A Systematic Review," *Proc. Des. Soc.*, vol. 2, pp. 1303–1312, May 2022, doi: 10.1017/pds.2022.132.

[29] M. W. Lauer-Schmaltz, P. Cash, and D. G. T. Rivera, "ETHICA: Designing Human Digital Twins—A Systematic Review and Proposed Methodology," *IEEE Access*, vol. 12, pp. 86947–86973, 2024, doi: 10.1109/access.2024.3416517.

[30] M. W. Lauer-Schmaltz, P. Cash, J. Paulin Hansen, and N. Das, "Human Digital Twins in Rehabilitation: A Case Study on Exoskeleton and Serious-Game-Based Stroke Rehabilitation Using the ETHICA Methodology," *IEEE Access*, vol. 12, pp. 180968–180991, 2024, doi: 10.1109/access.2024.3508029.

[31] N. Mandischer, A. Atanasyan, M. Schluse, J. Roßmann, and L. Mikelsons, "Perspectives-Observer-Transparency -- A Novel Paradigm for Modelling t he Human in Human-To-Anything Interaction Based on a Structured Review of the Human Digital Twin," *arXiv.org*, 2024, doi: 10.48550/ARXIV.2408.06785.

[32] A. Meghdari and M. Alemi, "Recent Advances in Social & Cognitive Robotics and Imminent Ethica l Challenges," in *Proceedings of the 10th International RAIS Conference on Social Scienc es and Humanities (RAIS 2018)*, Atlantis Press, 2018. doi: 10.2991/rais-18.2018.12.

[33] C. Montag and S. Diefenbach, "Towards Homo Digitalis: Important Research Issues for Psychology and t he Neurosciences at the Dawn of the Internet of Things and the Digital Society," *Sustainability*, vol. 10, no. 2, p. 415, Feb. 2018, doi: 10.3390/su10020415.

[34] D. Nguyen *et al.*, "Exploratory Models of Human-AI Teams: Leveraging Human Digital Twins t o Investigate Trust Development," *arXiv.org*, 2024, doi: 10.48550/ARXIV.2411.01049.

[35] C. Palmer, Y. M. Goh, E.-M. Hubbard, R. Grant, and R. Houghton, "The Need for a Symbiotic Interface for a Digital Twin," in *Advances in Transdisciplinary Engineering*, IOS Press, 2023. doi: 10.3233/atde230685.

[36] E. O. Popa, M. van Hilten, E. Oosterkamp, and M.-J. Bogaardt, "The use of digital twins in healthcare: socio-ethical benefits and soc io-ethical risks," *Life Sci Soc Policy*, vol. 17, no. 1, Jul. 2021, doi: 10.1186/s40504-021-00113-x.

[37] Y. (Wolf) Song, "Human Digital Twin, the Development and Impact on Design," *Journal of Computing and Information Science in Engineering*, vol. 23, no. 6, Aug. 2023, doi: 10.1115/1.4063132.

[38] E. Vildjiounaite *et al.*, "Challenges of learning human digital twin: case study of mental wellbe ing," in *Proceedings of the 16th International Conference on PErvasive Technolo gies Related to Assistive Environments*, ACM, Jul. 2023, pp. 574–583. doi: 10.1145/3594806.3596538.

[39] B. Wang *et al.*, "Human Digital Twin in the context of Industry 5.0," *Robotics and Computer-Integrated Manufacturing*, vol. 85, p. 102626, Feb. 2024, doi: 10.1016/j.rcim.2023.102626.

[40] E. W. R. and K. Ruben, "Keep IT Real: On Tools, Emotion, Cognition and Intentionality in Desig n," 2016.

[41] M. D. Xames and T. G. Topcu, "Toward Digital Twins for Human-in-the-loop Systems: A Framework for Wo rkload Management and Burnout Prevention in Healthcare Systems," in *2023 IEEE 3rd International Conference on Digital Twins and Parallel I ntelligence (DTPI)*, IEEE, Nov. 2023, pp. 1–6. doi: 10.1109/dtpi59677.2023.10365449.

Journal of Intelligent System and Applied Data Science (JISADS)

Journal homepage : https://www.jisads.com

ISSN (2974-9840) Online

BLOCKCHAIN BEYOND TECHNOLOGY: EXPLORING APPLICATIONS, COLLABORATIVE INNOVATIONS, AND GOVERNANCE STRATEGIES

Ihab Adib[1] and Youjun Liu[1]*

College of Life Science and Bio-Engineering, Beijing University of Technology, No. 100 Pingleyuan, Chaoyang District, Beijing 100124, China.

Ihab.s.adib@gmail.com, lyjlma@bjut.edu.cn

ABSTRACT

Blockchain is a new distributed computing paradigm characterized by security and trust, widely applied in various fields. However, security issues have become increasingly prominent, and the need for regulation is more urgent. The current state of the blockchain ecosystem and the regulatory policy backgrounds of major countries were briefly introduced. The relevant literature based on blockchain technology and application architecture were categorized and the characteristics of existing regulatory technologies and solutions were analyzed from three aspects: intra-chain regulation, inter-chain regulation, and off-chain regulation. Intra-chain regulation was further divided into three levels: infrastructure layer regulation, core function layer regulation, and user layer regulation. The advantages and disadvantages of different regulatory technologies at each level were discussed in detail. Inter-chain regulation was divided into two categories: regulation based on the "governance by chain" concept and cross-chain security regulation, with a brief discussion of the characteristics of related studies. Then some representative cases of off-chain regulation were introduced. Finally, the common issues in current blockchain security regulation were analyzed with possible improvement directions and new areas in need of regulation. The gap was filled in reviews on blockchain regulation and a reference for the design of blockchain regulatory solutions was provided.

Keywords: blockchain; blockchain security; blockchain regulation, Classification

1. INTRODUCTION

Since its inception, blockchain has evolved from Blockchain 1.0 to Blockchain 3.0, and its application scope has expanded from single payment scenarios to multiple industries, such as financial services, government and legal affairs, supply chain management, and identity verification [1]. Blockchain 1.0 focused on digital currencies, achieving decentralized value transfer. Blockchain 2.0 introduced smart contracts, marking the realization of complex business logic execution on-chain. Blockchain 3.0 emphasizes applying blockchain to real-world scenarios, realizing decentralized commercial networks [2].

In recent years, the rapid development of blockchain has led to increasingly rich blockchain applications. A batch of emerging blockchain projects represented by high-performance public chains has emerged, such as Solana [3], Avax [4], Near, Hedera [5], Sui [6], etc. Traditional public chains (such as Bitcoin, Ethereum, Binance Chain, etc.) have also attracted a large influx of funds, incubating various Web3 projects, such as decentralized exchanges (DEX) [7], decentralized social and chat software [8], inscription and rune protocols, blockchain games [9], Web3 cloud services [10], etc. [11-16].

With the explosive growth of blockchain technology applications, its security issues have also become prominent. The risks caused by vulnerabilities in underlying blockchain platforms and blockchain applications, as well as various virtual asset crimes, pose

great challenges to blockchain security. According to SlowMist Hacked Statistical database, the number of major public security incidents in global blockchain has shown an increasing trend year by year since 2012, as shown in Figure 1. Blockchain-related security incidents mainly include 9 categories: wallet security incidents, malicious mining, distributed denial-of-service (DDoS) attacks, ransomware, digital currency fraud, digital currency money laundering, smart contract security, exchange security, and other attack incidents [17-18].

With frequent blockchain security incidents, the demand for strengthening blockchain regulation is becoming increasingly urgent. Since 2019, although the number of blockchain-related literature included in databases such as IEEE, ACM, and Springer has reached more than 74,000, there are very few reviews directly studying blockchain regulation. Currently, domestic and foreign reviews related to blockchain regulation [19-22] tend to focus on the analysis of blockchain security or vulnerability detection and defense, or some related literature analyzes blockchain security in certain specific application scenarios [23-29], but does not involve blockchain regulation.

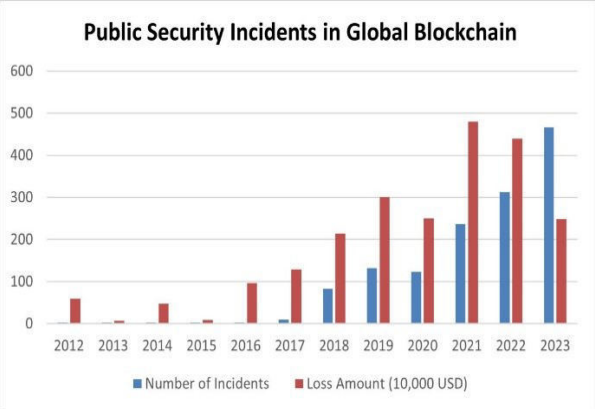

Figure1. Puplic security incedents in Global Blockchain

Based on the current development status of blockchain technology architecture, this paper divides it and the applications running on it into three layers: intra-chain infrastructure, cross-chain expansion, and decentralized autonomous communities and applications. The main contributions of this paper are as follows:

1) The existing regulatory schemes are summarized into intra-chain regulation, inter-chain regulation, and off-chain regulation. Intra-chain regulation is further divided into three levels: infrastructure layer regulation, core function layer regulation, and user layer regulation, and the advantages and disadvantages of regulatory technologies at each level are meticulously classified according to the focus of relevant literature.

2) Inter-chain regulation is further divided into regulation based on the "governance by chain" concept and cross-chain security regulation, analyzing and comparing the characteristics of related literature, and briefly discussing representative cases of off-chain regulation.

3) Common issues in existing regulation are analyzed, and possible improvement directions are provided, pointing out that regulation should focus on emerging blockchain projects represented by Rollup and decentralized finance (DeFi) projects.

1. Blockchain Regulation Background

With the in-depth development of blockchain technology, its application scenarios have gradually enriched, and various complex applications have gradually formed the embryonic form of the blockchain ecosystem. These ecological projects have attracted a continuous influx of massive funds, and at the same time, they have also attracted the attention of governments and organizations around the world. This section briefly introduces the current state of the blockchain ecosystem and the representative blockchain regulatory policies of major countries.

1.1 Current State of the Blockchain Ecosystem

In academia, some literature has proposed the concept of blockchain ecosystem [24-30,145]. After summarizing relevant literature [31-33], the composition of the blockchain ecosystem is shown in Figure 2. The bottom layer of Figure 3 is the supporting development technology, and breakthroughs often promote the innovation of blockchain technology, usually computer basic disciplines or technologies such as cryptography, big data, distributed systems, cloud and fog computing, and decentralized learning. The top-level application areas include real-world assets (RWA), electronic auctions, lending, decentralized finance, and many other scenarios. The blockchain ecosystem entities consist of eight parts: blockchain users, blockchain application providers, blockchain platform service providers, blockchain infrastructure, blockchain communities, blockchain equipment providers, blockchain regulatory agencies, and blockchain technology consulting providers [34-35]. These components are interconnected and interact through data and funds, forming an interdependent and mutually influential whole.

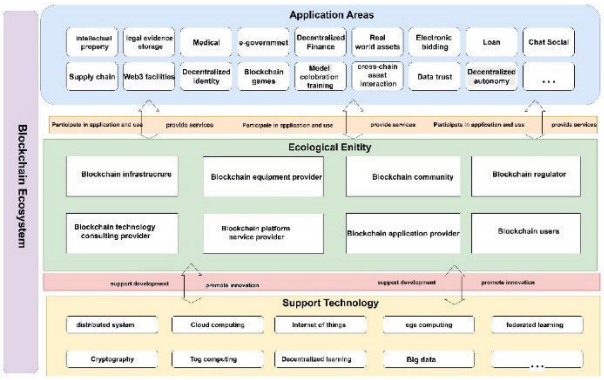

Figure 2. Blockchain ecosystem composition

1.2 Blockchain Regulatory Policies Beyond academic research, blockchain has received varying degrees of attention from governments and organizations worldwide during its development. Some countries and organizations have already carried out systematic and standardized blockchain security regulation work [36-37], and relevant regulatory policies and regulatory agencies are shown in Table 1.

Table 1: Blockchain-related regulatory policies and regulatory agencies of some countries and organizations

Canada	Canadian Cryptocurrency Tax Guide	Canada Revenue Agency (CRA)
France	Regulatory framework related to crypto assets and licensing requirements and regulations for digital asset service providers	French Financial Supervisory Authority (AMF); Association for the Development of Digital Assets (ADAN)
Singapore	Fiatech Regulatory Standard Guidelines, Digital Asset Taxation Act and Financial Institutions Code of Conduct	Monetary Authority of Singapore (MAS); Personal Data Protection Commission (PPDC); Intellectual Property Office of Singapore (IPOS)
China	Notice on Preventing Bitcoin Risks, Announcement on Preventing Token Issuance and Financing Risks, Blockchain Security White Paper, Blockchain Information Service Management Regulations and Financial Information Service Management Regulations	Cyberspace Administration of China; Digital Currency Research Institute of the People's Bank of China; Blockchain Committee of Internet Society of China (SEC)
Japan	Digital Finance Strategy, Blockchain Strategy, General Data Protection Regulation and EU Fintech Action Plan	European Securities and Markets Authority (ESMA); European Banking Authority (EBA); European Data Protection Supervisor (EDPS)
Germany	Digital Currency Exchange Act, Payment Services Act, Proposal on New ECO Regulation and Asset Settlement Act Enforcement Decree	Financial Services Agency of Japan (FAS), Japan Blockchain Association (LIRA), Japan Virtual Currency Exchange Association (JVCEA)
IMF	Crypto-assets, regulatory challenges in the global economy and regulatory frameworks for digital financial services. International Monetary Fund (IMF)	Federal Financial Supervisory Authority (BaFin); Federal Commissioner for Data Protection and Freedom of Information (BFDI); Financial Market Stability Fund (SoFFin)

The UK government first proposed the concept of a "regulatory sandbox" [38]. The United States has formulated intellectual property and tax regulations for blockchain technology and digital assets and established a blockchain industry alliance to promote the development and regulation of the blockchain industry [39]. The European Commission has formulated the "Digital Finance Strategy 2020" and "Blockchain Strategy" to strengthen regulation and cooperation in the digital finance field. The Singapore government has issued a digital asset tax law, stipulating that digital asset transactions should be taxed [40]. Switzerland has formulated a series of blockchain laws and policies [41] to provide legal protection, guidance, and regulation for blockchain enterprises. At the same time, it has also strengthened support and regulation of the blockchain industry, encouraging enterprises to develop more secure blockchain technology. In 2019, the Ministry of Industry and Information Technology established the National Blockchain and Distributed Ledger Technology Standardization Committee to systematically promote standardization work and accelerate the establishment of a blockchain regulatory system.

2. BLOCKCHAIN REGULATION AND CLASSIFICATION

Research progress on blockchain security and regulation at home and abroad is shown in Table 2. Literature [20-22, 45-49] focuses on the research of blockchain data security and network security issues, and does not discuss the overall regulation of blockchain. Literature [23-25, 50-56] focuses on discussing blockchain-specific security issues such as smart contract vulnerabilities and consensus algorithm vulnerabilities. In addition to the reviews listed in Table 2, there is also literature discussing the development of the blockchain ecosystem [26-30, 57], but it does not discuss blockchain security regulation.

This paper divides existing blockchain regulatory schemes and literature into intra-chain regulation, inter-chain regulation, and off-chain regulation. Intra-chain regulation consists of blockchain infrastructure layer regulation, core function layer regulation, and user layer regulation. It involves a large number of literatures [58-118] and is a key level of regulation. Inter-chain regulation consists of two types: regulation based on the "governance by chain" concept and cross-chain security regulation [119-131]. Off-chain regulation mainly involves decentralized autonomous organizations (DAO) and communities. Due to the short development history of cross-chain technology and off-chain decentralized governance mechanisms, there are fewer related literatures and regulatory schemes [132-137].

Table 2: Research progress on blockchain security and regulation

Literature	**Focus**	**Does it involve supervision?**
References[20-22,45]	Blockchain security, uncertainty, status of additional safeguards analyzed	Not involved
References[46-48]	Blockchain data security, network security, etc.	Not comprehensive
Literature[49]	Blockchain application research	Not involved
References[50-53]	Smart contract vulnerability detection and repair	Not involved
References[23-25,55-56]	Constraints mechanism security improvements	Not involved

2.1 Intra-chain Regulation

This section divides intra-chain regulation into three layers: infrastructure layer regulation, core function layer regulation, and user layer regulation. The regulatory technologies at the infrastructure layer are further divided into node association tracking, node abnormal behavior detection, and node attack traffic detection. The regulatory technologies at the core function layer are divided into abnormal transaction analysis and detection, smart contract security detection, consensus mechanism attack detection, and consortium chain penetration regulation. User layer regulation mainly targets users, including user business regulation and user account regulation.

2.1.1 Infrastructure Layer Regulation The infrastructure layer provides the necessary hardware components and operating environment to support the normal operation of the entire blockchain system, mainly including computing resources for storing blockchain data and executing blockchain computing tasks, backup and recovery mechanisms, and other security and protection measures to ensure the connectivity and data transmission stability of the network infrastructure between nodes.

2.1.1.1 Node Association Tracking Technology Blockchain node tracking technology refers to collecting and analyzing information such as network addresses, account addresses, and transactions of nodes in the blockchain network to construct the association relationship and topological structure between nodes, thereby understanding the connection methods, interaction situations, and transaction behavior characteristics between nodes, and achieving security regulation of the blockchain. Node association tracking technology does not affect the final state of transactions and blockchain, and belongs to ex-post regulation.

Related research [58-60] mainly uses graph analysis and log analysis, machine learning, and cluster analysis to track blockchain transactions, ultimately clarifying the relationship between blockchain nodes. The current difficulty is tracking highly private cryptocurrency transactions.

Graph analysis and log analysis audit. In response to the limitations of tracking analysis technology based on pollution/dyeing mentioned in literature [58-60] in terms of effectiveness, universality, and efficiency, Li Zhiyuan et al. [61] proposed a blockchain transaction tracking method based on node influence account balance model, which uses network analysis and graph data mining technology to track the flow of funds of specific target accounts through the account balance model, compensating for the shortcomings and deficiencies of existing blockchain transaction tracking research in terms of universality and efficiency. Focusing on the process tracking of consensus transactions, Li Shanshan et al. [62] proposed a Fabric consensus transaction trajectory tracking method based on custom logs, which uses the ELK (Elasticsearch Logstash Kibana) tool chain to collect and parse Fabric's custom consensus transaction logs, and processes custom log business logic through a Spring Boot backend application, which can effectively track the call trajectory of Fabric's consensus transactions at each node, realizing the visualization of consensus transaction trajectories. Focusing on node automatic discovery, literature [63] proposed a node automatic discovery mechanism based on the Kademlia protocol. The constructed routing table allows nodes in the network to gradually join their routing tables when discovered by other nodes, thereby realizing dynamic perception of the entire network by nodes.

Machine learning and cluster analysis. Using machine learning methods for blockchain node tracking can improve the efficiency and accuracy of tracking. Machine learning models can learn patterns and rules from a large amount of data, helping to identify and analyze complex node behaviors and relationships. Michalski et al. [64] used supervised learning methods to analyze the characteristics of nodes in the blockchain. By analyzing the behavioral characteristics of nodes in the blockchain network, they inferred the roles played by nodes in the blockchain, such as miners or exchanges. Although the goal of this paper is more focused on locating the roles and behaviors of nodes, its results can provide some help and clues for node tracking. Forward transaction tracking is a common technology used to analyze Bitcoin abuse and track fund flows, that is, starting from a given set of seed addresses known to belong to cybercrime activities, tracking the movement of Bitcoin, but it only considers forward transaction flows, and does not consider backward transaction flows, which means that in some cases, some important relationships and transaction information may be missed. In order to focus on both output transactions and input transactions when analyzing transactions, Gomez et al. [65] proposed a bidirectional exploration automated Bitcoin transaction tracking technology, which outputs a transaction graph from a given seed address belonging to cybercrime activities, and identifies the relationship paths between node activities and external services and other cybercrime activities. In order to prevent the transaction graph from expanding, this technology combines a labeled database with a machine learning classifier to quickly identify and filter out addresses belonging to exchanges. From the perspective of link prediction between nodes, Du et al. [66] proposed a graph neural network framework MixBroker, which uses original Ethereum mixed transaction data to construct a mixed currency interaction graph, and extracts account node features from the graph from multiple perspectives to better represent the attributes of mixed currency account nodes. The graph neural network is used to calculate the correlation probability between nodes, thereby determining the association relationship between mixed currency account nodes, which to a certain extent breaks the anonymity of Ethereum mixed currency services.

In addition, in order to provide a higher level of privacy and anonymity protection, some cryptocurrencies use ring signatures, zero-knowledge proofs, coin mixing, and other technical means to hide the addresses of both parties to the transaction and the transaction amount, such as Monero [67], Zcash [68], and Dash. Although these anonymous coins can provide a certain degree of anonymity and privacy protection, they are not untraceable. There are currently 6 types of Zcash [68] tracking technologies: Danan gift attack, dust attack, remote side-channel attack, round-trip transaction attack, user behavior analysis attack, and covert channel attack, which can be used to infer and track the transaction information of Zcash nodes. In the field of Monero [67] technology tracking, there are currently four main types of tracking methods: tracking based on input-output relationships [69] (such as 0-mixin attack, output merging attack, closed set attack, etc.), tracking based on statistical laws (such as latest guess attack, etc.), tracking with partially known public keys (such as flooding attack, wallet ring attack, etc.), and tracking using Monero's security mechanism vulnerabilities (such as malicious remote node attack, etc.).

2.1.1.2 Node Abnormal Behavior Detection Technology Blockchain nodes may attempt to perform malicious operations, attack networks, phish, tamper with data, or engage in fraudulent activities. Node detection refers to analyzing and monitoring node behavior in the blockchain network to identify possible malicious nodes or abnormal behaviors, and belongs to ex-ante regulation. Node detection methods are diverse, and currently mainly focus on traffic analysis and phishing node detection.

In terms of traffic analysis, Liu Guozhi [70] proposed an abnormal traffic detection algorithm based on federated learning and representation learning, and implemented a distributed abnormal traffic detection system for detecting abnormal nodes in the blockchain network. This system can automatically learn traffic data features, allow participants to dynamically enter and exit, and control the entire process through smart contracts. Unlike literature [70], which focuses on abnormal detection through specific algorithms and systems, Sanda et al. [71] used deep learning convolutional neural networks (CNN), K-nearest neighbors (KNN), decision trees, and multi-layer perceptrons (MLP) algorithms to determine classifiers and predict malicious nodes, which can be further extended to analyze abnormal behavior of verification nodes in proof of stake (PoS) consensus.

In terms of phishing node detection, current methods for detecting Ethereum network phishing mainly focus on transaction features and local network structure, but have limitations in handling complex interactions between edges and large graphs. In response to this problem, Zhang et al. [72] proposed an Ethereum phishing node detection method based on graph convolutional networks (GCN), which converts complex transaction networks into three simple inter-node graphs, and uses GCN to generate node embeddings and global structural information to identify phishing nodes. Similarly, Yu et al. [73] used a message-passing based GCN to first construct a transaction network, and then extract and classify node information to detect phishing nodes. Both of these works use GCN to detect Ethereum phishing nodes, and both involve the processing of transaction networks and the use of node information, which solves the limitations of current detection methods in handling complex interactions and large graphs, and improves the effectiveness and accuracy of detection. However, the former mainly focuses on using GCN to generate node embeddings and global structural information to identify

phishing nodes, while the latter focuses on first constructing a transaction network, and then extracting and classifying node information.

By timely identifying and responding to malicious nodes, abnormal behaviors, and potential risks, the anti-attack capability of blockchain systems can be enhanced, and a more reliable infrastructure can be provided for various application scenarios. However, node detection technology still faces some challenges, such as insufficient privacy protection during detection, low detection efficiency, and low accuracy.

2.1.1.3 Node Attack Traffic Detection Technology At the infrastructure layer, attacks that significantly harm the normal operation of blockchain nodes include Eclipse attacks [74] and DDoS attacks [76], whose purpose is to destroy the availability and functionality of the underlying network infrastructure. Researchers have proposed various detection methods based on deep learning to extract attack features from traffic data, focusing on how to identify and defend against attacks on blockchain infrastructure to ensure its stable and secure operation.

Eclipse Attack Detection Eclipse attacks rely on the cooperation of multiple nodes. By controlling the network connections of target nodes, the target nodes are isolated from other honest nodes. The client cannot distinguish between the canonical view of the blockchain and the view provided by the attacker. This attack has the characteristics of concealment and concurrency. Currently, most existing methods use custom behavior features and deep learning [74], immunity-based abnormal detection methods [77], suspicious timestamp-based detection methods, and communication using blockchain clients [78] to detect Eclipse attacks. In order to more accurately describe the behavioral characteristics of attack traffic, Dai et al. [74] enhanced the detection capability for Eclipse attack traffic by defining multi-level traffic features, improving the upsampling algorithm, and combining deep learning models, using CNN and bidirectional long short-term memory (Bi-LSTM) networks to extract deep features from Eclipse attack traffic, and integrating the feature extraction results into hybrid features through a multi-head attention mechanism. Detection based on suspicious block timestamps refers to determining whether the network is segmented by detecting the time interval between newly created blocks, but this method requires about 2-3 hours to relatively confirm whether the client is under attack. In order to reduce the average attack detection time, Alangot et al. [78] proposed that Bitcoin clients pass messages by establishing connections with servers on the Internet to exchange their blockchain views, and this method does not require introducing any dedicated infrastructure or changing the Bitcoin protocol and network.

Erebus Attack Detection Erebus attacks mainly target blockchain systems that use proof of work (PoW) consensus. Attackers interfere with the normal operation of target nodes by controlling a large number of IP addresses to form a fake network. In response to the problems of single detection objects, weak dynamic attack target perception, and high node resource requirements in existing methods, Dai et al. [75] designed a two-stage feature selection algorithm based on ReliefF_WMRmR and a multi-modal classification detection model based on deep learning by combining traffic behavior features with routing states based on multi-modal deep feature learning, and constructed a multi-modal neural network based on MLP, which can effectively detect Erebus attacks with high accuracy.

DDoS Attack Detection In terms of DDoS attack detection, Dai et al. [76] combined statistical and machine learning methods. By capturing traffic data at the node end of the blockchain network, cross-layer convolution operations are performed on the pre-processed traffic to extract abstract features of highly robust attack traffic, and an improved stochastic gradient descent algorithm is used to globally optimize the model parameters to prevent training parameter oscillation. Link flooding attack (LFA) is a new type of DDoS attack that uses low-rate traffic to flood a part of target links in the blockchain network to block normal traffic passing through these links and cut off the connection between the server and the network. In response to LFA, literature [79] used the time series prediction capability of long short-term memory networks to detect LFA, but whether it can accurately identify suspicious attack sources by calculating the similarity of different traffic sources remains to be further verified.

In addition, the visualization services and tools inherent in blockchain can be used as auxiliary tools for node association tracking and attack traffic detection. For example, blockchain browsers and data analysis tools such as Gephi, Cytoscape, Tokenview, and BlockAPI clearly present the transaction relationships or data interaction relationships between nodes or accounts.

In summary, for infrastructure layer regulation, blockchain node association tracking and detection technologies are mainly divided into two categories: one is to track their activities by monitoring message passing and transaction broadcasting between nodes in the blockchain network. Regulators can collect and analyze these data to understand node behavior patterns, network topology, and transaction flow; the other is to use data visualization technology to dynamically perceive the entire blockchain network through the routing table in blockchain nodes. Regulators can intuitively observe the connection relationships between nodes, transaction flows, and data changes. For the former, abnormal behaviors with defined detection rules can achieve relatively ideal results through data analysis. However, once new abnormal behaviors occur, new transaction datasets need to be organized and detection algorithms need to be redesigned for calculation, which has poor adaptability. The latter relies on data synchronization,

and it must be ensured that each node can achieve data consistency at a certain time. By dynamically visualizing operations to construct knowledge graphs, abnormal address clusters or nodes can be clearly discovered, making it easier to regulate these address clusters or nodes.

Overall, researchers tend to combine multiple technical means, especially graph analysis and machine learning, to achieve more intelligent node tracking and visualization at the infrastructure layer to improve the efficiency of blockchain regulation and the clarity of node relationships. Existing research explores how to improve the efficiency and accuracy of blockchain node tracking and detection. Some research focuses on specific tracking technologies and visualization methods, such as improving the efficiency of node tracking based on graph data mining, machine learning, and other technologies. Other research explores technical means to detect and defend against malicious nodes in different types of attacks (such as Eclipse attacks, DDoS attacks). The common goal of these studies is to enhance the security and regulability of blockchain networks, forming a multi-level, comprehensive node tracking and visualization framework. Future research directions may focus on improving the universality and efficiency of tracking technologies, exploring more secure and private tracking technologies, and further enhancing the security and robustness of blockchain networks.

2.1.2 Core Function Layer Regulation

The core function layer usually consists of core components such as transaction storage, transaction processing, and smart contracts, which are used to implement the basic functions and characteristics of the blockchain and provide reliable basic support for the user layer. The main regulatory methods are abnormal transaction analysis and detection, smart contract security detection, and consensus mechanism attack detection. In addition, consortium chains can achieve penetration regulation at the core function layer.

2.1.2.1 Abnormal Transaction Analysis and Detection Core function layer regulation mainly focuses on transaction data on the blockchain and the execution of smart contracts. Researchers have proposed various data analysis methods to analyze and detect on-chain data. A common method is to identify abnormal transactions and potential fraudulent behaviors based on data mining and machine learning technologies. Regulators can build models and algorithms to analyze the patterns and rules of transaction data and identify transaction behaviors that do not comply with the rules. Another method is to use graph theory and neural networks to analyze and study transaction flows and connection relationships in the blockchain network. By constructing transaction graphs and network maps, visualizing the relationships and connections between on-chain transaction data, identifying transaction flows, interaction patterns between addresses, and fund flow paths, abnormal nodes, transaction paths, and centralization can be discovered, thereby evaluating the security and stability of the blockchain network.

Abnormal Transaction Analysis and Detection Based on Data Mining and Machine Learning Currently, research on abnormal transaction analysis based on data mining and machine learning mainly focuses on deeply mining the features of blockchain node transaction data and discovering patterns and rules therein, so as to more effectively regulate the transaction behavior of blockchain networks. Zhu Huijuan et al. [80] proposed a blockchain abnormal transaction detection model, which adopts a residual network structure ResNet-32, and uses adaptive feature fusion methods to fully exploit the advantages of high-level abstract features and original features, improving the performance of blockchain abnormal transaction detection. This provides ideas for model construction and feature fusion for subsequent research. Taking the analysis of transaction motives as a starting point, Shen Meng et al. [81] designed a blockchain digital currency abnormal transaction behavior identification method based on motive analysis, selected airdrop candy and greedy funding as typical abnormal transaction behaviors, formulated judgment rules respectively, and abstracted abnormal transaction pattern diagrams, providing a reference for the classification and pattern research of abnormal transaction behaviors. Similarly, Zhang Xiaoqi et al. [82] proposed a network representation learning model DeepWalk-Ba for feature extraction of blockchain abnormal transactions. By constructing address and entity transaction graphs, combining features and machine learning for transaction entity identification, and extracting multi-granularity transaction patterns and user portraits based on transaction data analysis, timely and reliable detection of abnormal transactions in the blockchain can be achieved.

Abnormal Transaction Analysis and Detection Based on Graph Analysis and Neural Networks Wu et al. [83] designed two different community detection methods for Bitcoin and Ethereum networks, respectively proposing specific clustering algorithms derived from spectral clustering algorithms and novel community detection algorithms for low-level signals on graphs, helping to find user communities based on user token subscriptions. Further, Lin Wei [84] studied abnormal transaction data detection based on blockchain technology, and proposed a blockchain abnormal transaction data detection model based on a custom sliding window mechanism, a fully connected neural network, and a multi-channel output feature vector fusion of bidirectional gated recurrent units. In order to protect user privacy and reduce the risk of data being illegally obtained or abused during detection, Chen Binjie et al. [85] proposed a KNN-based blockchain abnormal transaction detection scheme with privacy protection. Accounting nodes randomize transaction data features by using matrix multiplication,

and then the cloud server uses KNN to detect abnormal transactions on the randomized transaction data features.

In terms of abnormal detection of smart contracts in blockchain, Liu et al. [86] proposed detecting fraudulent contracts by using transaction data and code data of Ethereum smart contracts, extracting features from complex smart contracts, effectively identifying abnormal contracts, and constructing a heterogeneous graph transformation network suitable for abnormal detection of smart contracts to detect financial fraud. However, whether more precise feature extraction methods can be developed to improve the efficiency of smart contract abnormal detection still needs further in-depth exploration.

2.1.2.2 Smart Contract Security Detection Smart contracts, as a core component of blockchain technology, have received much attention for their security issues. Research in this area is currently relatively mature [53, 87-90]. To ensure brevity, this section only briefly discusses relevant literature from a regulatory perspective.

Smart contract vulnerability detection methods include static analysis, dynamic analysis, formal verification, metamorphic testing, and graph neural network-based methods. These methods aim to identify potential vulnerabilities in smart contracts, such as reentrancy attacks, integer overflows, permission issues, and timestamp dependency issues. Since abnormal detection of smart contracts occurs before or after transactions and does not affect the final transaction results, this belongs to ex-ante or ex-post regulation.

1) Static Analysis By statically scanning and analyzing contract code, potential vulnerabilities are detected. Common tools include SmartCheck, Slither, etc.

2) Dynamic Analysis By simulating contract execution and monitoring its behavior, potential security issues can be found. ReGuard [91] generates random and diverse transaction data using fuzz testing, simulates possible attack scenarios, and dynamically identifies potential reentrancy attacks in smart contracts by recording key execution traces.

3) Formal Verification Verifies whether smart contracts comply with expected design attributes and security specifications. ZEUS [92] is an automated formal verification tool for smart contracts, which converts Solidity source code into LLVM (low-level virtual machine) intermediate language, and uses XACML (eXtensible access control markup language) to design five security vulnerability detection rules to determine the security of target programs during formal verification.

4) Metamorphic Testing By generating test cases and executing them in smart contracts, it verifies whether the test results meet expectations. In response to possible security vulnerabilities, Chen Jinfu et al. [93] designed different metamorphic relationships and performed metamorphic testing. By verifying whether the source test cases and subsequent test cases satisfy the metamorphic relationship, it determines whether there are related security vulnerabilities in the smart contract.

5) Deep Learning Based on the source code, operation code, and control flow patterns of smart contracts, features are extracted, and deep learning models (such as CNN, RNN, and Transformer) are used to train and predict whether there are security vulnerabilities. Deng et al. [94] proposed a smart contract vulnerability detection method using deep learning and multi-modal decision fusion, considering the code semantics and control structure information of smart contracts, and integrating source code, operation code, and control flow patterns through multi-modal decision fusion. Zhang et al. [95] proposed a hybrid deep learning model - convolutional and bidirectional gated recurrent unit (CBGRU), which combines word embedding methods (Word2Vec, FastText) and deep learning methods (LSTM, GRU, Bi-LSTM, CNN, BiGRU). Word embedding methods can convert words or phrases into vector representations to capture their semantic relationships. Different deep learning models extract smart contract features from different perspectives, combine them, and input them into a classifier for smart contract vulnerability detection.

Smart contract security is an important and complex field in blockchain technology. Many studies have been devoted to the detection and repair of smart contract security, but most vulnerability detection tools can only detect single and old versions of smart contract vulnerabilities [96]. Future research should focus on further improving the automation, efficiency, and accuracy of detection tools, combining static analysis methods with dynamic analysis methods to detect more types of vulnerabilities in multi-version smart contracts, thereby achieving higher detection accuracy.

2.1.2.3 Consensus Mechanism Attack Detection Consensus protocols are sets of rules in blockchain systems that determine transaction verification and block addition. Some common and harmful attacks include double-spending attacks, 51% attacks, selfish mining attacks, and saving attacks. Research on 51% attacks and double-spending attacks is relatively extensive and mature [97-100]. To ensure brevity, only saving attacks and selfish mining attacks, which have a greater impact on regulation, are briefly discussed.

Saving Attack is a new type of attack that can delay nodes from reaching consensus. Attackers "save" their proposed blocks during temporary consensus failures and use these rights to trigger another consensus failure after the network returns to normal, which leads to a decrease in blockchain performance and an increase in the delay of block finalization. Otsuki et al. [101] conducted a simulation study of Saving Attack on various fork selection rules, including the longest chain rule, GHOST (greedy heaviest-observed sub-tree), LMD GHOST (latest-message-driven GHOST), and FMD GHOST (fresh-message-driven GHOST). The research results show that Saving Attack has a very negative impact on consensus. Under experimental conditions, an attacker with 30% voting power successfully prevented LMD GHOST consensus for 83 minutes after saving their blocks for 32 minutes.

Selfish mining attacks are carried out by a small number of malicious miners or mining pools who exploit vulnerabilities or potential weaknesses in the system design to obtain more mining rewards unfairly. Wang et al. [102] used machine learning methods to detect selfish mining attacks in blockchain. They used logistic regression and fully connected neural networks (including 10 hidden layers and 10 neurons per layer) to train classification models on the training set, and judged whether unknown samples belonged to selfish mining attacks by learning the features of the samples, or belonged to ex-post regulation methods.

2.1.2.4 Consortium Chain Penetration Regulation Consortium chain penetration regulation is mainly located at the core function layer. Penetration regulation is a method introduced from the financial field into blockchain regulation, which refers to the regulation and traceability of all nodes and transaction data on the consortium chain through the penetration of regulatory nodes to ensure the security and stable operation of the consortium chain, and belongs to in-process regulation methods. In consortium chain regulation, regulatory logic can be embedded in the components of the core function layer, so penetration regulation can go deep into each entity for regulation and supervise and audit all transactions and information.

Liu Huixia et al. [103] proposed a blockchain-based security regulation scheme for shared charging piles, constructing a shared charging trust model based on a dual chain. They built a trust relationship between transaction parties through authentication contracts and designed a penetration regulation scheme to verify the identity of users, pile owners, or operators upwards, and verify the accuracy of charging amount, charging speed, and other information downwards, effectively regulating all participants and specific transaction data of car shared charging. Wang et al. [105] proposed an illegal data hierarchical interception scheme based on consortium chains. By using regular expressions and smart contracts at the application end to mark and block illegal data with different degrees of impact, it can effectively regulate illegal data in the blockchain. Different from previous single consortium chains, Zhang Jianyi et al. [106] adopted a regulable digital currency model with a consortium chain-public chain dual-chain structure, which uses the consortium chain as the core participant in consensus, ensures the privacy of user transaction data through secret sharing, and at the same time uses the public chain as the operating basis, allowing ordinary users to participate in and witness the maintenance of the system. In order to achieve comprehensive protection of transaction privacy and fine-grained mandatory regulation, Huo Xinlei et al. [107] proposed a consortium chain scheme with authorized regulation and privacy protection functions, including the division of member roles under the consortium chain and chameleon hash functions, zero-knowledge proofs, and other cryptographic technologies. Literature [106] focuses on the dual-chain structure and user participation, while literature [107] focuses on achieving comprehensive and fine-grained regulation and privacy protection through technical means.

In multiple application scenarios of consortium chains, researchers have also proposed some personalized solutions. In the field of agricultural machinery scheduling, Yang et al. [108] proposed a consortium blockchain-based agricultural machinery scheduling system. The upper-layer regulation improves the efficiency and security of the consensus algorithm and allows supervisors to block users with malicious intentions, ensuring the security of the system and improving the transparency and efficiency of data flow in the field of agricultural machinery scheduling. In the field of construction engineering, Li et al. [109] proposed the TABS (two-layer adaptive blockchain-based supervision) model for supervising off-site modular housing production, which realizes communication and transactions between adaptive private side chains and the main chain, ensuring the authenticity of operation records and protecting participant privacy, providing an unalterable and privacy-preserving regulatory mechanism for the construction engineering industry.

In addition, regulatory agencies can be considered as privileged nodes to access the consortium chain, and the effect of penetration regulation can be achieved by tracing and auditing on-chain data, which is a feasible regulatory direction.

In summary, core function layer regulation, in terms of abnormal transaction detection, researchers have proposed various methods to detect and analyze abnormal transactions on the blockchain, including abnormal transaction identification methods based on data mining and machine learning technologies, and using graph theory and neural networks to analyze and study transaction flows and connection relationships. Different studies have proposed various models and algorithms. For example, Zhu Huijuan et al. [80] used a residual network structure to improve detection

performance, while Zhang Xiaoqi et al. [82] performed transaction entity identification through network representation learning. Although these methods differ in technical details, their common goal is to improve the regulatory capabilities of blockchain networks and ensure the legality of transaction behavior. In terms of smart contract security detection, it can be seen that researchers refer to and learn from each other's work in abnormal transaction analysis and smart contract security. The static analysis, dynamic analysis, and formal verification methods mentioned in the literature complement each other and detect potential vulnerabilities in smart contracts from different angles. For example, formal verification methods verify whether smart contracts comply with design attributes, while dynamic analysis methods discover security issues by simulating smart contract execution. These studies are jointly committed to improving the security of smart contracts and reducing the risks brought by potential vulnerabilities. In terms of consortium chain regulation, in contrast to public chains, since regulation can be introduced as a basic function into the core function layer, or regulatory parties can be connected as nodes with regulatory authority, consortium chains can achieve penetration regulation.

2.1.3 User Layer Regulation

The user layer provides blockchain interfaces, blockchain nodes, user wallets, and other functions, supporting developers and miners to participate, use, and maintain the blockchain.

2.1.3.1 User Business Regulation User business regulation at the user layer mainly focuses on user business aspects, such as double-spending, false transactions, money laundering, Ponzi schemes, illegal token issuance, etc. Abnormal transaction behavior analysis and detection methods can be used to detect such businesses. Abnormal transaction behavior refers to the behavior of participants in a blockchain system that does not conform to normal transaction behavior patterns. In response to these abnormal transaction behaviors, the design and regulatory mechanisms of blockchain systems need to consider security and compliance, including identifying abnormal transaction behaviors, monitoring transaction patterns, and implementing compliance rules, etc., which are ex-post regulation methods, to reduce and prevent the occurrence of abnormal transaction behaviors. Related research focuses on abnormal transaction behaviors of blockchain users and corresponding regulatory mechanisms, which correspond to blockchain users and blockchain regulatory agencies, respectively.

Currently, blockchain abnormal transaction behavior identification methods have problems such as unclear identification targets, low efficiency in processing massive data, and single identification dimensions. In response to these problems, Zhao Zening [110] proposed an incremental identification method based on heuristic address clustering and a transaction behavior prediction method based on transaction subgraph partitioning, which improved the address clustering algorithm and improved the prediction accuracy by constructing transaction graphs and using graph neural networks. Qu Yuan [111] studied abnormal transaction behaviors in Bitcoin from two levels: macroscopic traffic data and microscopic transaction data. For macroscopic traffic data, unsupervised abnormal analysis and alarm functions were achieved by combining support vector machines and encoders and decoders. For microscopic transaction data, evolutionary graph convolutional networks (GCN) and time graph attention (TGA) mechanisms were used for feature extraction, and random forests were used for abnormal detection and alarming of illegal transactions, providing a more comprehensive abnormal detection solution. Existing solutions have improved identification accuracy, detection precision, and efficiency, but whether machine learning algorithms and encryption technologies can be combined to enhance the existing blockchain abnormal transaction behavior identification effect and privacy protection function still needs further in-depth exploration.

2.1.3.2 User Account Regulation Private keys are crucial for users to access their accounts and assets. Hackers may attack users' wallets, obtain private keys or tamper with transaction information by forging identities, inducing or deceiving users, thereby stealing assets. Ethereum has attracted a large number of users and developers, however, malicious users and attackers also use the anonymity and openness of Ethereum to engage in various illegal activities, such as pyramid schemes, fraud, money laundering, etc. Researchers have proposed machine learning, graph analysis, and time series analysis methods for Ethereum accounts to detect and identify malicious accounts, which belong to ex-ante or ex-post regulation methods.

In response to transaction security issues caused by fraudulent accounts in blockchain, Zhou Jian et al. [112] proposed a fraudulent account detection and feature analysis model based on machine learning, and introduced SHAP values to provide a more accurate prediction model through on-chain data feature analysis. Farrugia et al. [113] proposed a new method for detecting illicit users in Ethereum, which detects illicit activities on the Ethereum network at the account level by feature extraction and feature importance analysis, combined with the XGBoost classification model.

Liang Fei et al. [114-115] successively proposed methods based on hyperbolic graph neural convolutional networks (LSC-GCN) [114] and subspace graph clustering (GCN-Clustering) [115] to detect malicious Ethereum accounts. In response to the problems of insufficient labels in datasets leading to insufficient model training and low identification efficiency in existing models, GCN-Clustering converts original node address features into node features containing cluster

information, uses the clustering information of the dataset itself to enhance the feature extraction capability of nodes, and at the same time uses GCN for supervised learning, further strengthening the embedding expression of cluster information obtained in unsupervised learning in node features.

Shi Tuo et al. [116] incorporated transaction time information into the Ethereum address account feature model, proposed a graph attention mechanism based on time series transaction relationships, and improved the traditional attention network. By using the attention mechanism, the central node and neighboring nodes are aggregated, which can effectively identify Ethereum addresses with abnormal transaction behaviors.

For the wallet security of Bitcoin and three privacy-focused cryptocurrencies: Dash, Monero, and Zcash, Biryukov et al. [117] manually checked and used static analysis tools (such as FlowDroid, SmartDec Scanner) to scan and analyze wallets, detecting security threats in wallet installation methods, permission requirements, and privacy policies. They proposed a transaction clustering method based on transaction time analysis, listening to network traffic and attempting to associate attackers' cryptocurrency addresses with IP addresses or other identity information.

In summary, for the identification and regulation of abnormal behaviors (such as double-spending, false transactions, money laundering) in blockchain systems, researchers have proposed a series of methods based on heuristic address clustering, transaction subgraph partitioning, Ethereum Ponzi scheme detection, etc., aiming to improve the accuracy and efficiency of abnormal behavior identification. In addition, this section also focuses on account security and regulation issues, especially the theft of assets due to private key leakage. Existing research uses machine learning, graph analysis, and time series analysis methods to detect and identify malicious accounts and improve account security. With the continuous evolution of blockchain technology and the expansion of its application scope, future research can develop towards more refined abnormal behavior detection methods, more effective account security protection strategies, and more in-depth data analysis and mining technologies to adapt to increasingly complex and diverse security threats. At the same time, with the continuous improvement and strengthening of regulatory regulations, researchers also need to pay more attention to the compliance of blockchain systems to ensure their sustainable development and widespread application in business and finance.

Table 3 shows the comparison of blockchain regulatory technologies related to the infrastructure layer, core function layer, and user layer, where × indicates that it is not considered or is not the focus of the solution, and √ indicates that the solution is involved.

2.2 Inter-chain Regulation

Inter-chain regulation focuses on the interaction and interoperability regulation between different blockchains. The core services of inter-chain regulation are cross-chain asset exchange, inter-chain communication and data sharing, cross-chain App operations, smart contract interoperability, decentralized identity authentication, etc. There are two types of inter-chain regulation: one is to deploy regulatory logic on the regulatory chain based on the core idea of "governance by chain," where the regulated chain synchronizes data with the regulatory chain, and the regulatory chain can operate on the regulated chain; the other is to regulate existing cross-chain protocol

Table 3. Comparison of blockchain regulatory technologies

Section	Blockchain Monitoring Technology Comparison	Applicable	Monitoring Method	Network Security
Supervision Layer	Use ELK + Kafka + Fabric for transaction data trading monitoring	Fabric	Strong	✗
	Based on Kademlia protocol for node autonomous discovery	Public	Strong	✗
	Based on Ethereum/chain of blocks and the concept of account-based models of blockchain	Bitcoin	Strong	✗
	Node capability-based account structure model for blockchain transmission and method	Ethereum	Public	✗
	Use machine learning to predict abnormal node behavior based on network data	Bitcoin	Strong	✗
	Double spending attack detection based on Bitcoin trading data analysis (124)	Bitcoin	Strong	✗
Basic Infrastructure	Detection of nodes in the network and classify them, analyzing malicious nodes' behavior	Public	Medium	✓
	Ability to classify nodes based on node interaction behavior	Fabric	Weak	✗

	Based on GNX node classification, detect network nodes (123-125)	Ethereum	Strong	✗
	Use based on anomaly detection methods to study network nodes (126-127)	Ethereum	Strong	✗
	Use of KNN for network node classification	Public	Medium	✓
	Use of random forest for abnormal node detection based on the network environment	Ethereum	Strong	✗
Node Behavior Analysis & Detection Methods	Detect and analyze new abnormal behavior of nodes based on current network data (128)	Public	Medium	✗
	Based on signature and machine learning to detect new abnormal behaviors, such as DDoS attacks (134)	Public	Medium	✗
	Use of graph algorithms to analyze communication relationships between nodes (134)	Public	Medium	✗
	Use of deep learning to classify malicious behaviors, such as KNN for abnormal behavior detection (91)	Public	Medium	✓
Core Trust & Security	Automatic detection of trustworthiness based on behavior model (138)	Public	Medium	✗
	Construct local security models based on communication behaviors (92)	Public/Weak	Weak	✗
	Use abnormal behavior models for detection and prevention of malicious behaviors (90)	Public	Medium	✗
	Detect new malicious behavior patterns in blockchain communication (91)	Public	Medium	✗
Shared Mechanisms for Security	Use of simulations to analyze the impact of malicious behaviors on consensus (101)	POS	Strong	✗
	Use of replay attacks and attack simulation models to analyze malicious attacks (102)	Bitcoin	Strong	✗
	Construct-based security mechanism for shared liability and design trust protocols (103)	Ethereum	Strong	✗
	Role-based permission control based on security policies for secure access (104)	Ethereum	Medium	✓
Distributed Identity and User Management	Use of identity management and identity authentication based on blockchain (105)	POS	Strong	✗
	Use of blockchain for user identity management and privacy protection (106)	Bitcoin	Medium	✓
	User behavior tracking and account analysis based on user activity (107)	Ethereum	Weak	✗
	Use of LSC-GCN for GCN-Clustering methods to analyze user behaviors (113-116)	Ethereum	Strong	✗

2.2.1 Regulation Based on the "Governance by Chain" Concept Kevin Werbach et al. [119] first proposed the concept of "governance by chain" in the legal field. Chen Chun [57] further deepened this concept. The basic principle of "governance by chain" technology is to use one blockchain as a regulatory chain to regulate another blockchain, i.e., the regulated chain. The regulator can create a smart contract on the regulatory chain, which stipulates the rules and conditions to be complied with on the regulated chain. This leads to an important research direction - blockchain "compliance" regulation, which aims to ensure that blockchain transactions and activities comply with legal regulations, norms, and standards. These requirements can be any type of rule, such as transaction restrictions, prevention of double-spending, anti-money laundering, etc. When some nodes or users on the regulated chain violate these rules, the regulator can initiate sanctions on the regulatory chain through smart contracts. These sanctions usually involve penalties or disciplinary measures, such as freezing accounts, prohibiting transactions, or revoking transactions.

Ethereum, through ERC (Ethereum Request for Comments), standardizes smart contracts. From ERC20 to ERC1400, it has achieved a shift from avoiding regulation to embracing regulation [120]. ERC20 only requires providing functions such as token issuance and transfer, while ERC1400 stipulates the standard for issuing security tokens, requiring smart contracts to provide relevant legal documents and perform restriction judgments before executing transfers, providing readable explanations of judgment results, thereby realizing functions such as locking positions at the contract level, KYC/AML verification, and freezing in/out accounts. Libra also released White Paper 2.0 in 2020 to respond to regulatory concerns, including compliance controls (such as VASP certification, non-custodial wallet restrictions, etc.), making all transactions on the Libra blockchain enforce certain

compliance requirements. These measures are all aimed at improving the compliance and transparency of blockchain transactions and better adapting to regulatory requirements. Boya Zheng Chain provides a smart contract programming language RegLang [121] for regulatory technology. According to regulatory needs, it designs the syntax rules and type system of contracts. Regulators can automatically implement penetration regulation through smart contracts. Regulated objects can improve automated compliance capabilities through regulatory rules published by regulators, improving regulatory efficiency and accuracy, and making regulation more standardized, intelligent, and digital. Lu et al. [122] built the OriginChain system to provide transparent, tamper-proof, and traceable data, and automatically perform compliance checks. The system generates smart contracts representing legal agreements, automatically checks and executes services and terms, and checks whether legal and regulatory requirements are met. Mao Xiangke et al. [123] built a blockchain system with regulatory functions and rollback operations, realizing regulation of blockchain transactions at three different stages: pre-event, in-event, and post-event.

Some domestic enterprises are also actively promoting the implementation of "governance by chain" technology. Tencent Security released the "CCGP Cross-Chain Governance White Paper," realizing "governance by chain" cross-chain interoperability and collaboration. This system has five major advantages: strong universality, easy scalability, multi-party co-governance, high efficiency, high security, and traceable records, covering three application scenarios: wide-area data sharing, joint traceability, and wide-area evidence storage, which is expected to promote the application of blockchain technology in multiple scenarios. The Beijing Internet Court issued the "Tianping Chain" application access technology and management specifications [124], which standardize the technology and process of blockchain application access, improving the credibility and efficiency of electronic evidence. This specification involves three aspects: system security of the access platform, compliance of electronic data, and security of blockchain, promoting the application of blockchain technology in the judicial field. Literature [125-128] discusses smart contract compliance verification models in different application scenarios such as IoT, law, and cloud services, verifying and confirming the compliance of smart contracts in different environments.

Jing Pujie et al. [129] proposed a hierarchical cross-chain regulatory architecture based on the idea of "governance by chain," and designed a "regulatory chain-business chain" cross-chain collaborative governance model in the regulatory architecture, which improved the centralized and authoritarian nature of regulatory behavior. The designed cross-chain interaction standard data structure with universality ensures the smooth, secure, and efficient cross-chain regulatory process. Zhang et al. [130] proposed their on-chain hierarchical structure, on-chain and off-chain hybrid storage model, on-chain regulatory process, and traceable transaction information process. Through pre-event, in-event, and post-event collaborative regulation, multi-party hierarchical and multi-dimensional regulation of the entire data transaction process is achieved, and regulatory smart contracts are used to achieve hierarchical regulation of multiple regulators and post-event traceability (ex-post regulation), which can effectively isolate and protect sensitive information between data transactions.

2.2.2 Cross-chain Security Regulation Cross-chain technology is an important technical means to achieve inter-chain interconnection and value transfer. Cross-chain technology realizes interoperability and data exchange between different blockchains, but it also brings new security risks.

The security of cross-chain systems mainly depends on atomicity, inter-chain information synchronization, and network channel security. Given the diversity of heterogeneous blockchains in terms of block structure, consensus mechanisms, and complex working mechanisms, coupled with inherent security vulnerabilities in cross-chain technology, such as defects in the principles and implementation mechanisms of cross-chain technology, all these factors may cause security risks. In addition, if the consensus algorithm of the underlying blockchain has vulnerabilities or is compromised, the security of cross-chain interactive operations will also be threatened.

The notary mechanism may lead to collusion attacks and single point of failure risks. Notaries are nodes responsible for verifying and confirming cross-chain transactions. If notaries collude or a notary is attacked, the security of the entire cross-chain system will be threatened. The hash lock mechanism is a time-constrained mechanism used for cross-chain transactions, which may be affected by clock drift and malicious delay attacks. Clock drift may lead to inaccurate lock times, while malicious delay attacks exploit network delays to manipulate the execution order of cross-chain transactions. Wu Di [131] proposed defense methods against hash lock transfer delay attacks, relay cross-chain routing attacks, and relay chain block blocking attacks, which to a certain extent strengthened the security regulation of cross-chain systems. First, to prevent hash lock transfer delay attacks, the time difference can be increased. By increasing the time difference between the Fabric end and the ETH end, the difficulty for attackers to maliciously wait and block the network can be increased. Then, three protection methods can be adopted to deal with relay cross-chain routing attacks: application chain whitelist, application chain balance query, and application chain creation time query. Finally, by comprehensively using two methods: setting connection count scripts and modifying the

gateway's request processing order, relay chain block blocking attacks can be effectively prevented.

2.3 Off-chain Regulation

Off-chain regulation refers to regulators regulating and managing regulated chains through off-chain mechanisms, including community discussions, voting, negotiations, off-chain governance, committee decisions, and other methods. However, off-chain regulation has problems such as insufficient participation, abuse of power, and lack of transparency [132-133], which need to be solved through effective mechanisms and rules.

The Ethereum The DAO incident [134] and the Bitcoin block size debate [135] are two typical off-chain regulatory events. The Ethereum The DAO incident involved the security and governance issues of Ethereum smart contracts. Finally, the Ethereum community decided to hard fork the Ethereum blockchain through off-chain discussions and voting to recover stolen assets and maintain the stability of the Ethereum network. The Bitcoin block size debate lasted for several years, involving important matters such as Bitcoin network protocol updates and capacity expansion. However, the final decision was made by a few developers and miners through off-chain negotiations and voting, and most Bitcoin users did not participate in or understand this process. This lack of transparency and insufficient participation in regulation reflects some problems and limitations of off-chain regulation, and also triggers discussions and attempts at off-chain regulation. For example, the block node election protocol Whisk proposed by the Ethereum Open Research Forum Ethresearch was discussed and designed by multiple community members rather than official Ethereum personnel.

In practice, a combination of on-chain and off-chain regulation can achieve better regulatory and community governance effects. EOS [136] is a blockchain project based on the delegated proof of stake (DPoS) consensus algorithm, and its community governance mechanism adopts a combination of on-chain and off-chain regulation. Off-chain regulation includes community discussions, voting, and negotiations, while on-chain regulation is implemented through smart contracts. Miyachi et al. [137] proposed a modular hybrid privacy-preserving framework for enhancing medical information management, combining on-chain and off-chain regulation to design a reference model. It mainly realizes the interaction between on-chain and off-chain resources through a distributed software architecture, thereby realizing privacy management of different types of medical data.

3. FUTURE OUTLOOK OF BLOCKCHAIN REGULATION

From the analysis and summary of the three categories of blockchain regulatory technologies in Section 3, it can be seen that there are four common problems in current blockchain regulation.

1. Difficulty in Data Association Analysis Blockchain transaction data is stored in a distributed network. Due to the decentralization and anonymity of blockchain transactions, it is difficult for regulators to track the true identity of transaction participants. For example, on privacy public chains such as Monero, Dash, and Zcash, the identities of transaction participants and transaction details are not public, making it difficult for regulators to obtain complete transaction information, thereby making it difficult to discover and punish violations. It is difficult to regulate illegal transactions and behaviors in these blockchain networks.

A possible solution is to break through the association of chain group entities and anonymous digital identity recognition technologies, build a three-in-one associated regulation of blockchain entities-data-chain groups, and integrate machine learning to extract features of non-anonymous data such as network layer traffic data, and train targeted regulatory large language models. However, the security of the unique algorithms of large language models in blockchain security regulation also needs to be considered to ensure the security of the regulatory technology itself. A typical attack method against large language models is command injection. Attackers can construct inputs cleverly to make the model perform unexpected behaviors. If the blockchain regulatory interface based on large language models is abused, even with input specifications, attackers may still use command injection to exploit the authority of the regulatory interface, causing damage or interfering with the normal operation of the regulated blockchain application.

2.Insufficient Consideration of Business Compliance Regulation Existing regulatory schemes tend to use technical means to regulate a specific vulnerability or risk, ignoring the compliance and security risks of the regulatory target business itself, which may lead to regulatory loopholes. Existing regulatory methods and technologies [80-81, 84, 113] are generally less versatile. It should be considered to regulate on-chain business and security vulnerability risks collaboratively, and design specialized regulatory schemes or systems for business and technical risks respectively.

3.Low Cross-chain Collaboration Regulation Capability Blockchain cross-chain protocols have matured, and various cross-chain projects have emerged. Cross-chain is no longer limited to involving only two blockchains, but has evolved into complex cross-chain scenarios with multi-chain collaborative interconnection represented by

Polkadot. In this regard, corresponding blockchain regulatory research is not yet deep and sufficient. It is necessary to consider establishing cross-chain regulatory interoperability mechanisms [139] or multi-chain collaborative regulatory mechanisms, such as using Polkadot's parachain auction mechanism to embed regulatory logic into the obtained parachains, and regulating blockchain applications connected to the parachains.

4.High Regulatory Cost Since the operation of regulatory schemes or systems requires continuous external investment of resources, regulatory costs will only increase, and it is impossible to achieve self-sustaining regulation. For example, node detection and attack detection technologies require long-term maintenance of necessary network facilities or deployment of nodes to collect blockchain P2P layer traffic. Abnormal transaction analysis and smart contract security require a large amount of computing resources to train the necessary machine learning models to complete detection or identification. Node tracking technology requires a large amount of data analysis. Penetration regulation requires a large amount of software resources to meet regulatory requirements.

A possible way to balance regulatory costs is for blockchain regulators, as members of the blockchain community, to propose and vote on matters as members of decentralized autonomous organizations. The benefits generated by these processes can be used to reduce regulatory costs. Therefore, whether it is possible to quantify and model regulatory effectiveness and regulatory benefits using game theory based on the blockchain ecosystem model and regulatory costs, thereby further analyzing the specific role of regulation in the development of the blockchain ecosystem, is a direction that needs to be explored.

With the in-depth development of blockchain technology, various Rollup [140] projects aimed at solving the scalability problems of existing public chains have emerged, such as Arbitrum [141], Optimism [142], etc., as well as high-performance public chains adopting new accounting structures or sharding, such as Kaspa [143], Near [144], etc. The applicability of traditional regulatory technologies to them needs to be further tested. In addition, the emergence of decentralized exchanges has promoted the prosperity of the decentralized finance ecosystem, and the regulation of decentralized exchanges will be a key area of blockchain security regulation.

For the regulation of these emerging blockchain projects, feasible regulatory measures are as follows:

A. Regulation should consider using decentralized autonomous organizations to achieve regulation. For example, the decentralized communities of permissionless chains themselves have governance rights and voting rights for projects. These communities have low participation thresholds and are a major effective way of regulation.

B. The scope of regulation should be extended to various Rollup solutions and DeFi projects, and targeted regulation should be carried out according to their underlying implementation mechanisms, thereby increasing the coverage of regulation.

C. Attention should be paid to the new Bitcoin ecosystem and targeted regulation should be carried out. Recently, inscription ecosystems represented by Ordinals and Sats, rune ecosystems represented by Runes, and Bitcoin smart contract virtual machines have emerged. In the future, regulators should pay attention to these emerging blockchain projects.

4. CONCLUSION

The rapid development of blockchain has brought increasingly serious security issues, making blockchain security regulation a key research area. This paper analyzes and summarizes the current state of the blockchain ecosystem and briefly explains the domestic and international policy background of blockchain regulation. Based on the characteristics of current blockchain technology and its applications, it provides a three-layer division of intra-chain infrastructure, cross-chain expansion, and off-chain decentralized autonomous communities and applications. Based on this division, existing regulatory technologies and schemes are summarized and systematically analyzed and compared from three aspects: intra-chain regulation, inter-chain regulation, and off-chain regulation. The paper focuses on discussing relevant literature on infrastructure layer, core function layer, and user layer regulation within intra-chain regulation and compares their characteristics. It briefly discusses representative schemes for inter-chain and off-chain regulation, and finally summarizes and compares the three regulatory schemes: intra-chain, inter-chain, and off-chain. It also points out common problems in current blockchain security regulation, possible improvement directions, and emerging blockchain projects that regulators should pay attention to in the future.

REFERENCES

[1] CHOI T M, SIQIN T. Blockchain in logistics and production from blockchain 1.0 to blockchain 5.0: an intra-inter-organizational framework[J]. Transportation Research Part E: Logistics and Transportation Review, 2022, 160: 102653.

[2] ALIEF R N, PUTRA M A P, GOHIL A, et al. FLB2: layer 2 blockchain implementation scheme on federated learning technique[C]//Proceedings of the 2023 International Conference on Artificial Intelligence in Information and Communication (ICAIIC). Piscataway: IEEE Press, 2023: 846-850.

[3] PIERRO G A, TONELLI R. Can solana be the solution to the blockchain scalability problem? [C]//Proceedings of the 2022 IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER). Piscataway: IEEE Press, 2022: 1219-1226.

[4] ROCKET T, YIN M F, SEKNIQI K, et al. Scalable and probabilistic leaderless BFT consensus through metastability[J]. arXiv Preprint, arXiv: 1906.08936, 2019.

[5] TANG Y, YAN J W, CHAKRABORTY C, et al. Hedera: a permissionless and scalable hybrid blockchain consensus algorithm in multiaccess edge computing for IoT[J]. IEEE Internet of Things Journal, 2023, 10(24): 21187-21202.

[6] FITZI M, WANG X C, KANNAN S, et al. Minotaur: multi-resource blockchain consensus[C]//Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM Press, 2022: 1095-1108.

[7] JAYAPAL C, M J, S N R. An insight into NFTs, stablecoins and DEXs in blockchain[C]//Proceedings of the 2023 2nd International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA). Piscataway: IEEE Press, 2023: 1-6.

[8] DEVMANE M A. D-space: a decentralized social media app[C]//Proceedings of the 2023 2nd International Conference on Edge Computing and Applications (ICECAA). Piscataway: IEEE Press, 2023: 809-814.

[9] BREIKI H A. Trust evolution game in blockchain[C]//Proceedings of the 2022 IEEE/ACS 19th International Conference on Computer Systems and Applications (AICCSA). Piscataway: IEEE Press, 2022: 1-4.

[10] KARANJAI R, XU L, DIALLO N, et al. DeFaaS: decentralized function-as-a-service for emerging dApps and Web3[C]//Proceedings of the 2023 IEEE International Conference on Blockchain and Cryptocurrency (ICBC). Piscataway: IEEE Press, 2023: 1-3.

[11] GABRIEL T, CORNEL-CRISTIAN A, ARHIP-CALIN M, et al. Cloud storage. A comparison between centralized solutions versus decentralized cloud storage solutions using blockchain technology[C]//Proceedings of the 2019 54th International Universities Power Engineering Conference (UPEC). Piscataway: IEEE Press, 2019: 1-5.

[12] ZHONG Z S, WEI S R, XU Y T, et al. SilkViser: a visual explorer of blockchain-based cryptocurrency transaction data[C]//Proceedings of the 2020 IEEE Conference on Visual Analytics Science and Technology (VAST). Piscataway: IEEE Press, 2020: 95-106.

[13] SABLE N P, RATHOD V U, SABLE R, et al. The secure E-wallet powered by blockchain and distributed ledger technology[C]//Proceedings of the 2022 IEEE Pune Section International Conference (PuneCon). Piscataway: IEEE Press, 2022: 1-5.

[14] ARAB G A, COGLIATTI J I, URQUIZÓ P, et al. Development of a blockchain-based Web3 application for CO2 absortion right management[C]//Proceedings of the 2023 IEEE International Humanitarian Technology Conference (IHTC). Piscataway: IEEE Press, 2023: 1-4.

[15] CUI W P, SUN Y X, ZHOU J R, et al. Understanding the blockchain ecosystem with analysis of decentralized applications: an empirical study[C]//Proceedings of the 2021 the 5th International Conference on Management Engineering, Software Engineering and Service Sciences. New York: ACM Press, 2021: 38-44.

[16] SUN J, SADDIK A E, CAI W. Smart contract as a service: a paradigm of reusing smart contract in web3 ecosystem[J]. IEEE Consumer Electronics Magazine, 2024, 14(1): 46-55..

[17] RAIKWAR M, GLIGOROSKI D. Aggregation in blockchain ecosystem[C]//Proceedings of the 2021 Eighth International Conference on Software Defined Systems (SDS). Piscataway: IEEE Press, 2021: 138-143.

[18] DONG W L, LIU Z, LIU K, et al. Survey on vulnerability detection technology of smart contracts[J]. Journal of Software, 2024, 35(1): 38-62.

[19] WEI S J, LÜ W L, LI S S. Overview on typical security problems in public blockchain applications[J]. Journal of Software, 2022, 33(1): 324-355.

[20] HAN X, YUAN Y, WANG F Y. Security problems on blockchain: the state of the art and future trends[J]. Acta Automatica Sinica, 2019, 45(1): 206-225.

[21] LIU M D, CHEN Z N, SHI Y J, et al. Research progress of blockchain in data security[J]. Chinese Journal of Computers, 2021, 44(1): 1-27.

[22] LIU A D, DU X H, WANG N, et al. Research progress on blockchain system security technology[J]. Chinese Journal of Computers, 2024, 47(3): 608-646.

[23] ZHOU S S, LI K, XIAO L J, et al. A systematic review of consensus mechanisms in blockchain[J]. Mathematics, 2023, 11(10): 2248.

[24] XU J, WANG C, JIA X H. A survey of blockchain consensus protocols[J]. ACM Computing Surveys, 2023, 55(13): 1-35.

[25] CHOO K R, OZCAN S, DEHGHANTANHA A, et al. Editorial: blockchain ecosystem: technological and management opportunities and challenges[J]. IEEE Transactions on Engineering Management, 2020, 67(4): 982-987.

[26] KHANG A, CHOWDHURY S, SHARMA S. The data-driven blockchain ecosystem: fundamentals, applications, and emerging technologies[M]. Boca Raton: CRC Press, 2022.

[27] RIASANOW T, BURCKHARDT F, SETZKE D S, et al. The generic blockchain ecosystem and its strategic implications[C]//Proceedings of the 24th Americas Conference of Information Systems. Piscataway: IEEE Press, 2018. 1-10.

[28] REHMAN M H U, SALAH K, DAMIANI E, et al. Trust in blockchain cryptocurrency ecosystem[J]. IEEE Transactions on Engineering Management, 2020, 67(4): 1196-1212.

[29] STAFFORD T F, TREIBLMAIER H. Characteristics of a blockchain ecosystem for secure and sharable electronic medical records[J]. IEEE Transactions on Engineering Management, 2020, 67(4): 1340-1362.

[30] KABASHKIN I. Risk modelling of blockchain ecosystem[C]//International Conference on Network and System Security. Berlin: Springer, 2017: 59-70.

[31] YOO S. A study on blockchain ecosystem[J]. The Journal of the Institute of Webcasting, Internet and Telecommunication, 2018, 18: 1-9.

[32] KIM J W. Analysis of blockchain ecosystem and suggestions for improvement[J]. Journal of Information and Communication Convergence Engineering, 2021, 19(1): 8-15.

[33] RAIKWAR M, GLIGOROSKI D. DoS attacks on blockchain ecosystem[C]//European Conference on Parallel Processing. Berlin: Springer, 2022: 230-242.

[34] ZHANG H, YI J B, WANG Q. Research on the collaborative evolution of blockchain industry ecosystems in terms of value co-creation[J]. Sustainability, 2021, 13(21): 11567.

[35] PAPADONIKOLAKI E, TEZEL A, YITMEN I, et al. Blockchain innovation ecosystems orchestration in construction[J]. Industrial Management & Data Systems, 2023, 123(2): 672-694.

[36] ZHANG W, DONG W, ZHANG F Q, et al. The application of German blockchain technology in the field of financial science and technology, its supervision ideas and its enlightenment to China[J]. International Finance, 2019(9): 76-80.

[37] YANG D, CHEN Z L. Virtual currency legislation: experience of Japan and inspiration to China[J]. Securities Market Herald, 2018(2): 69-78.

[39] DENG J P. Blockchain regulatory mechanism and enlightenment in United States[J]. China Policy Review, 2019(1): 125-130.

[40] DENG J P. Singapore's blockchain regulatory policy and its review[J]. Fudan University Law Review, 2020(1): 59-72.

[41] PI L Y, XUE Z W. Regulation arrangement and international practice of crypto-asset transactions[J]. Securities Market Herald, 2019(7): 4-12.

[45] LIU H Q, RUAN N. A survey on attacking strategies in blockchain[J]. Chinese Journal of Computers, 2021, 44(4): 786-805.

[46] YU G, NIE T Z, LI X H, et al. The challenge and prospect of distributed data management techniques in blockchain systems[J]. Chinese Journal of Computers, 2021, 44(1): 28-54.

[47] XU K, LING S T, LI Q, et al. Research progress of network security architecture and key technologies based on blockchain[J]. Chinese Journal of Computers, 2021, 44(1): 55-83.

[48] QIN C X, GUO B, SHEN Y, et al. Security risk assessment model of blockchain[J]. Acta Electronica Sinica, 2021, 49(1): 117-124.

[49] QIAN P, LIU Z G, HE Q M, et al. Smart contract vulnerability detection technique: a survey[J]. Journal of Software, 2022, 33(8): 3059-3085.

[50] CUI Z Q, YANG H W, CHEN X, et al. Research progress of security vulnerability detection of smart contracts[J]. Journal of Software, 2024, 35(5): 2235-2267.

[51] JIANG F, CHAO K L, XIAO J M, et al. Enhancing smart-contract security through machine learning: a survey of approaches and techniques[J]. Electronics, 2023, 12(9): 2046.

[52] WU H G, PENG Y B, HE Y Q, et al. A review of deep learning-based vulnerability detection tools for Ethernet smart contracts[J]. Computer Modeling in Engineering & Sciences, 2024, 140(1): 77-108.

[53] CHU H T, ZHANG P C, DONG H, et al. A survey on smart contract vulnerabilities: data sources, detection and repair[J]. Information and Software Technology, 2023, 159: 107221.

[54] CHEN J F, FENG Q W, CAI S H, et al. Vulnerability detection model for blockchain systems based on formal method[J]. Journal of Software, 2024, 35(9): 4193-4217.

[55] WANG Y, GOU G P, LIU C, et al. Survey of security supervision on blockchain from the perspective of technology[J]. Journal of Information Security and Applications, 2021, 60: 102859.

[56] YE C C, LI G Q, CAI H M, et al. Security detection model of blockchain[J]. Journal of Software, 2018, 29(5): 1348-1359.

[57] CHEN C. Key technologies of consortium blockchain and regulatory challenges of blockchain[R]. 2019.

[58] MÖSER M, BÖHME R, BREUKER D. Towards risk scoring of Bitcoin transactions[C]//Financial Cryptography and Data Security. Berlin: Springer, 2014: 16-32.

[59] ANDERSON R. Making Bitcoin legal (transcript of discussion) [C]//Security Protocols XXVI. Berlin: Springer, 2018: 254-265.

[60] TOVANICH N, CAZABET R. Pattern analysis of money flows in the Bitcoin blockchain[C]//International Conference on Complex Networks and Their Applications. Berlin: Springer, 2023: 443-455.

[61] LI Z Y, XU B L, ZHOU Y Y. Blockchain anonymous transaction tracking method based on node influence[J]. Computer Science, 2024, 51(7): 422-429.

[62] LI S S, WANG Y Z, ZOU Y L, et al. Consensus transaction trajectory visualization tracking method for Fabric based on custom logs[J]. Journal of Computer Applications, 2022, 42(11): 3421-3428.

[63] ZHENG L W, HELU X H, LI M H, et al. Automatic discovery mechanism of blockchain nodes based on the kademlia algorithm[C]//International Conference on Artificial Intelligence and Security. Berlin: Springer, 2019: 605-616.

[64] MICHALSKI R, DZIUBAŁTOWSKA D, MACEK P. Revealing the character of nodes in a blockchain with supervised learning[J]. IEEE Access, 2020, 8: 109639-109647.

[65] GOMEZ G, MORENO-SANCHEZ P, CABALLERO J. Watch your back: identifying cybercrime financial relationships in Bitcoin through back-and-forth exploration[C]//Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM Press, 2022: 1291-1305.

[66] DU H B, CHE Z, SHEN M, et al. Breaking the anonymity of ethereum mixing services using graph feature learning[J].

IEEE Transactions on Information Forensics and Security, 2024, 19: 616-631.

[67] LIN D K, YAN J Q, LANDENG B, et al. Survey of anonymity and tracking technology in monero[J]. Journal of Computer Applications, 2022, 42(1): 148-156.

[68] FU Z H, LIN D K, JIANG H C, et al. Survey of anonymous and tracking technology in zerocash[J]. Computer Science, 2021, 48(11): 62-71.

[69] KUMAR A, FISCHER C, TOPLE S, et al. A traceability analysis of monero's blockchain[C]//European Symposium on Research in Computer Security. Berlin: Springer, 2017: 153-173.

[70] LIU G Z. Research and implementation of abnormal traffic monitoring method based on federated learning[D]. Beijing: Beijing University of Posts and Telecommunications, 2021.

[71] SANDA O, PAVLIDIS M, SERAJ S, et al. Long-range attack detection on permissionless blockchains using deep learning[J]. Expert Systems with Applications, 2023, 218: 119606.

[72] ZHANG Z, HE T, CHEN K, et al. Phishing node detection in ethereum transaction network using graph convolutional networks[J]. Applied Sciences, 2023, 13(11): 6430.

[73] YU T, CHEN X M, XU Z, et al. MP-GCN: a phishing nodes detection approach via graph convolution network for ethereum[J]. Applied Sciences, 2022.

[74] DAI Q Y, ZHANG B, DONG S Q. Eclipse attack detection for blockchain network layer based on deep feature extraction[J]. Wireless Communications and Mobile Computing, 2022, 2022(1): 1451813.

[75] DAI Q Y, ZHANG B, XU K Y, et al. An Erebus attack detection method oriented to blockchain network layer[J]. Computers, Materials & Continua, 2023, 75(3): 5395-5431.

[76] DAI Q Y, ZHANG B, DONG S Q. A DDoS-attack detection method oriented to the blockchain network layer[J]. Security and Communication Networks, 2022, 2022: 5692820.

[77] LYU J S, YANG P, CHEN W, et al. Abnormal detection of eclipse attacks on blockchain based on immunity[J]. Computer Science, 2018, 45(2): 8-14.

[78] ALANGOT B, REIJSBERGEN D, VENUGOPALAN S, et al. Decentralized and lightweight approach to detect eclipse attacks on proof of work blockchains[J]. IEEE Transactions on Network and Service Management, 2021, 18(2): 1659-1672.

[79] CAO W(J/Q). Research on detection method of trusted link flooding attack based on blockchain[D]. Hefei: University of Science and Technology of China, 2022.

[80] ZHU H J, CHEN J F, LI Z Y, et al. Block-chain abnormal transaction detection method based on adaptive multi-feature fusion[J]. Journal on Communications, 2021, 42(5): 41-50.

[81] SHEN M, SANG A Q, ZHU L H, et al. Abnormal transaction behavior recognition based on motivation analysis in blockchain digital currency[J]. Chinese Journal of Computers, 2021, 44(1): 193-208.

[82] ZHANG X Q, BAI X, LI G S, et al. Blockchain abnormal transaction detection based on network representation learning[J]. Cyber Security and Data Governance, 2022, 41(10): 11-20.

[83] WU S X, WU Z X, CHEN S H, et al. Community detection in blockchain social networks[J]. Journal of Communications and Information Networks, 2021, 6(1): 59-71.

[84] LIN W. Detection of abnormal transactions in blockchain based on multi feature fusion[J]. Netinfo Security, 2022(10): 24-30.

[85] CHEN B J, WEI F S, GU C X. Blockchain abnormal transaction detection with privacy-preserving based on KNN[J]. Netinfo Security, 2022, 22(3): 78-84.

[86] LIU L, TSAI W T, BHUIYAN M Z A, et al. Blockchain-enabled fraud discovery through abnormal smart contract detection on Ethereum[J]. Future Generation Computer Systems, 2022, 128: 158-166.

[87] HE D J, DENG Z, ZHANG Y X, et al. Smart contract vulnerability analysis and security audit[J]. IEEE Network, 2020, 34(5): 276-282.

[88] VACCA A, SORBO A D, VISAGGIO C A, et al. A systematic literature review of blockchain and smart contract development: Techniques, tools, and open challenges[J]. Journal of Systems and Software, 2021, 174: 110891.

[89] CHEN H S, PENDLETON M, NJILLA L, et al. A survey on ethereum systems security: vulnerabilities, attacks, and defenses[J]. ACM Computing Surveys, 2020, 53(3): 1-43.

[90] KANNENGIEßER N, LINS S, SANDER C, et al. Challenges and common solutions in smart contract development[J]. IEEE Transactions on Software Engineering, 2022, 48(11): 4291-4318.

[91] LIU C, LIU H, CAO Z, et al. ReGuard: finding reentrancy bugs in smart contracts[C]//Proceedings of the 40th International Conference on Software Engineering: Companion Proceeedings. New York: ACM Press, 2018: 65-68.

[92] KALRA S, GOEL S, DHAWAN M, et al. ZEUS: analyzing safety of smart contracts[C]//Proceedings 2018 Network and Distributed System Security Symposium. Piscataway: IEEE Press, 2018: 1-12.

[93] CHEN J F, WANG Z X, CAI S H, et al. Vulnerability detection method for blockchain smart contracts based on metamorphic testing[J]. Journal on Communications, 2023, 44(10): 164-176.

[94] DENG W C, WEI H C, HUANG T, et al. Smart contract vulnerability detection based on deep learning and multimodal decision fusion[J]. Sensors, 2023, 23(16): 7246.

[95] ZHANG L J, CHEN W J, WANG W Z, et al. CBGRU: a detection method of smart contract vulnerability based on a hybrid model[J]. Sensors, 2022, 22(9): 3577.

[96] HE D J, WU R, LI X J, et al. Detection of vulnerabilities of blockchain smart contracts[J]. IEEE Internet of Things Journal, 2023, 10(14): 12178-12185.

[97] RAMEZAN G, LEUNG C. Analysis of proof-of-work-based blockchains under an adaptive double-spend attack[J]. IEEE Transactions on Industrial Informatics, 2020, 16(11): 7035-7045.

[98] ZHENG J, HUANG H W, ZHENG Z B, et al. Adaptive double-spending attacks on PoW-based blockchains[J]. IEEE Transactions on Dependable and Secure Computing, 2024, 21(3): 1098-1110.

[99] SAAD M, SPAULDING J, NJILLA L, et al. Exploring the attack surface of blockchain: a comprehensive survey[J]. IEEE Communications Surveys & Tutorials, 2020, 22(3): 1977-2008.

[100] CHULERTTIYAWONG D, JAMALIPOUR A. Sybil attack detection in Internet of flying things-IoFT: a machine learning approach[J]. IEEE Internet of Things Journal, 2023, 10(14): 12854-12866.

[101] OTSUKI K, NAKAMURA R, SHUDO K. Impact of saving attacks on blockchain consensus[J]. IEEE Access, 2021, 9: 133011-133022.

[102] WANG Z J, LV Q Z, LU Z B, et al. ForkDec: accurate detection for selfish mining attacks[J]. Security and Communication Networks, 2021, 2021(1): 5959698.

[103] LIU H X, LI L L. Security supervision scheme of shared charging pile based on blockchain[J]. Application Research of Computers, 2022, 39(5): 1319-1323, 1348.

[105] WANG X Q, ZHANG K, DING Y, et al. An illegal data supervision scheme for the consortium blockchain[C]//Blockchain Technology and Application. Berlin: Springer, 2022: 100-115.

[106] ZHANG J Y, WANG Z Q, XU Z L, et al. A regulatable digital currency model based on blockchain[J]. Journal of Computer Research and Development, 2018, 55(10): 2219-2232.

[107] HUO X L, LONG Y, GU D W. Privacy protection and authorization supervision scheme based on consortium chain[J]. Journal of Chinese Computer Systems, 2023, 44(3): 589-595.

[108] YANG H T, XIONG S M, FRIMPONG S A, et al. A consortium blockchain-based agricultural machinery scheduling system[J]. Sensors, 2020, 20(9): 2643.

[109] LI X, WU L, ZHAO R, et al. Two-layer adaptive blockchain-based supervision model for off-site modular housing production[J]. Computers in Industry, 2021, 128: 103437.

[110] ZHAO Z N. Research on key technologies of blockchain abnormal trading behavior identification[D]. Tianjin: Tianjin University of Technology, 2023.

[111] QU Y. Research and design of Bitcoin abnormal behavior detection system[D]. Chengdu: University of Electronic Science and Technology of China, 2021.

[112] ZHOU J, ZHANG J, YAN S. Research on blockchain fraud account detection based on data on chain[J]. Application Research of Computers, 2022, 39(4): 992-997.

[113] FARRUGIA S, ELLUL J, AZZOPARDI G. Detection of illicit accounts over the ethereum blockchain[J]. Expert Systems with Applications, 2020, 150: 113318.

[114] LIANG F, WEI L, LIN W C. A method for detecting malicious Ethereum accounts based on subspace graph clustering [J]. Journal of Information Security Research, 2023, 9(E1): 68-71.

[115] LIANG F, MA L, ZHAI B Y, et al. Detection of malicious accounts in ethereum based on hyperbolic space graph neural convolution network[J]. Civil-Military Integration on Cyberspace, 2022(9): 48-52.

[116] SHI T, LIANG F, SHANG G C, et al. Detection of malicious ethereum account based on time series transaction and graph attention neural network[J]. Netinfo Security, 2022, 22(10): 69-75.

[117] BIRYUKOV A, TIKHOMIROV S. Security and privacy of mobile wallet users in Bitcoin, Dash, Monero, and Zcash[J]. Pervasive and Mobile Computing, 2019, 59: 101030.

[118] XU W K. Node detection algorithm for preventing blockchain bifurcation[J]. Electronic Technology & Software Engineering, 2020(3): 186-187.

[119] WERBACH K, LIN S W. Trust, but verify: why the blockchain needs the law[J]. Oriental Law, 2018(4): 83-115.

[120] KONG H D. Study on the legal nature of non-homogeneous general certificate[J]. Network Security Technology & Application, 2022(9): 141-143.

[121] GAO J B, ZHANG J S, LI Q S, et al. RegLang: a smart contract programming language for regulation[J]. Computer Science, 2022, 49(6): 462-468.

[122] LU Q H, XU X W. Adaptable blockchain-based systems: a case study for product traceability[J]. IEEE Software, 2017, 34(6): 21-27.

[123] MAO X K, LI C, HAO Y T, et al. A blockchain system design and implementation for all-round supervision[J]. Computer & Digital Engineering, 2023, 51(1): 81-85, 92.

[124] ZHANG Y N, WU P C. The construction of “balance chain” in Beijing Internet court and its enlightenment: also on the feasibility of blockchain technology to maintain the authenticity of electronic files[J]. Archives & Construction, 2022(10): 63-65.

[125] AMATO F, COZZOLINO G, MOSCATO F, et al. A model for verification and validation of law compliance of smart contracts in IoT environment[J]. IEEE Transactions on Industrial Informatics, 2021, 17(11): 7752-7759.

[126] PARVIZIMOSAED A, SHARIFI S, AMYOT D, et al. Subcontracting, assignment, and substitution for legal contracts in symboleo[C]//Conceptual Modeling. Berlin: Springer, 2020: 271-285.

[127] MOLINA-JIMENEZ C, SFYRAKIS I, SOLAIMAN E, et al. Implementation of smart contracts using hybrid architectures with on and off-blockchain components[C]//Proceedings of the 2018 IEEE 8th International Symposium on Cloud and Service Computing (SC2). Piscataway: IEEE Press, 2018: 83-90.

[128] PARVIZIMOSAED A, BASHARI M, KIAN A R, et al. Compliance checking for transactive energy contracts using smart contracts[C]//Proceedings of the 2020 IEEE PES Transactive Energy Systems Conference (TESC). Piscataway: IEEE Press, 2020: 1-5.

[130] ZHANG Y Q, MA Z F, LUO S S, et al. DBSDS: a dual-blockchain security data sharing model with supervision and privacy-protection[J]. Concurrency and Computation: Practice and Experience, 2023, 35(21): e7706.

[131] WU D. Research on multi-scenario attack and defense method for cross-chain system[D]. Beijing: Beijing Jiaotong University, 2022.

[132] BRINKMANN M, HEINE M. Can blockchain leverage for new public governance: a conceptual analysis on process level[C]//Proceedings of the 12th International Conference on Theory and Practice of Electronic Governance. New York: ACM Press, 2019: 338-341.

[133] DURSUN T, ÜSTÜNDAĞ B B. A novel framework for policy based on-chain governance of blockchain networks[J]. Information Processing & Management, 2021, 58(4): 102556.

[134] DIROSE S, MANSOURI M. Comparison and analysis of governance mechanisms employed by blockchain-based distributed autonomous organizations[C]//Proceedings of the 2018 13th Annual Conference on System of Systems Engineering (SoSE). Piscataway: IEEE Press, 2018: 195-202.

[135] USHIDA R, ANGEL J. Regulatory considerations on centralized aspects of DeFi managed by DAOs[C]//Financial Cryptography and Data Security. Berlin: Springer, 2021: 21-36.

[136] MONCADA R, FERRO E, FAVENZA A, et al. Next generation blockchain-based financial services[C]//Euro-Par 2020: Parallel Processing Workshops. Berlin: Springer, 2021: 30-41.

[137] MIYACHI K, MACKEY T K. hOCBS: a privacy-preserving blockchain framework for healthcare data leveraging an on-chain and off-chain system design[J]. Information Processing & Management, 2021, 58(3): 102535.

[138] BARATI M, RANA O. Tracking GDPR compliance in cloud-based service delivery[J]. IEEE Transactions on Services Computing, 2022, 15(3): 1498-1511.

[139] YANG D. “Rule of law by chain” and “Chain-based governance”: the integration path of blockchain technology regulation [R]. 2019.

[140] GORZNY J, LIN P A, DERKA M. Ideal properties of rollup escape hatches[C]//Proceedings of the 3rd International Workshop on Distributed Infrastructure for the Common Good. New York: ACM Press, 2022: 7-12.

[141] KALODNER H A, GOLDFEDER S, CHEN X Q, et al. Arbitrum: scalable, private smart contracts[C]//27th USENIX Security Symposium. Berkeley: USENIX Association, 2018: 1353-1370.

[142] GONÇALVES J P D B, VILLAÇA R D S. A new consensus mechanism for blockchained federated learning systems using optimistic rollups[C]//Proceedings of the 2024 IEEE International Conference on Blockchain. Piscataway: IEEE Press, 2024: 406-411.

[143] SOMPOLINSKY Y, WYBORSKI S, ZOHAR A. PHANTOM GHOSTDAG: a scalable generalization of nakamoto consensus: September 2, 2021[C]//Proceedings of the 3rd ACM Conference on Advances in Financial Technologies. New York: ACM Press, 2021: 57-70.

[144] SNEHLATA, SHUKLA P, SINGH A K, et al. An intelligent blockchain-oriented digital voting system using NEAR protocol[J]. SN Computer Science, 2023, 4(5): 643.

[145] Aloun, M. S. (2024). Synergistic Integration of Artificial Intelligence and Blockchain Technology: Advancements, Applications, and Future Directions. Journal of Intelligent Systems and Applied Data Science, 2(2).

Journal of Intelligent System and Applied Data Science (JISADS)

Journal homepage : https://www.jisads.com

ISSN (2974-9840) Online

A COMPREHENSIVE REVIEW OF FEDERATED LEARNING: ADVANCEMENTS, CHALLENGES, AND FUTURE DIRECTIONS

Ahsan Wajahat [1], Kailong Zhang [1], Jahanzaib Latif [1],*

*[1] * School of Software, Northwestern Polytechnical University, Xi'an, 710129, China.*

ahsan.sunny56@yahoo.com , ki.zhang@nwpu.edu.cn, jahanzaib@nwpu.edu.cn

ABSTRACT

Federated Learning (FL) has emerged as a groundbreaking distributed machine learning paradigm that enables collaborative model training while preserving data privacy. This comprehensive review examines FL's evolution from its inception to current state-of-the-art approaches, addressing both theoretical foundations and practical applications. We analyze the core FL framework, highlighting its advantages over centralized learning in terms of privacy preservation, reduced communication overhead, and edge computing capabilities. The paper explores key algorithmic advancements including Federated Averaging (FedAvg) and its variants (FedProx, SCAFFOLD), which tackle challenges like data heterogeneity and client drift. We discuss FL's transformative applications across healthcare, finance, and IoT domains, where data privacy is paramount. Major challenges are critically examined, including communication bottlenecks, straggler effects, security vulnerabilities, and the complexities of non-IID data distributions. The review evaluates privacy-enhancing technologies such as differential privacy and homomorphic encryption, analyzing their trade-offs between privacy guarantees and model performance. Looking forward, we identify promising research directions: adaptive personalization techniques, integration with large language models, blockchain-assisted security frameworks, and standardization efforts for broader adoption. Ethical considerations and regulatory compliance aspects are also addressed, providing a holistic perspective on FL's role in shaping responsible AI development. This review serves as both a technical reference and a roadmap for future innovation in federated learning systems.

Keywords: Federated Learning, Privacy-Preserving AI, Edge Computing, Decentralized Optimization

1. INTRODUCTION

The proliferation of data in the modern digital landscape, coupled with an escalating global emphasis on data privacy, has presented traditional machine learning paradigms with formidable challenges. Conventional approaches often necessitate the aggregation of vast datasets in centralized repositories to train robust and accurate models. This centralized model, however, is increasingly constrained by stringent privacy regulations, such as GDPR and CCPA, which mandate strict control over personal data[22]. Furthermore, the centralized storage of massive datasets introduces inherent risks, including potential data breaches, corruption, loss, and significant storage and management overheads. These limitations underscore the urgent need for innovative machine learning methodologies that can circumvent the pitfalls of data centralization while still harnessing the collective intelligence embedded within distributed datasets.

In response to these pressing concerns, Federated Learning (FL) has emerged as a transformative paradigm. Conceived by Google in 2016, FL is a decentralized and collaborative machine learning approach that enables multiple entities to jointly train a shared global model without exchanging their raw, sensitive data. Instead of data moving to the computation, computation moves to the data. This fundamental shift ensures that private data remains localized on individual devices or institutional servers, thereby upholding stringent privacy standards and mitigating the risks associated with

centralized data collection. The core principle of FL lies in its iterative process: a central server orchestrates the training, distributing a global model to participating clients. Each client then trains this model on its local dataset, computes model updates (e.g., gradients or model parameters), and securely transmits only these updates back to the central server. The server then aggregates these updates to refine the global model, which is subsequently redistributed for another round of local training. This cycle continues until the model converges or a predefined performance threshold is met[1].

The advantages of federated learning extend beyond privacy preservation. It effectively addresses the challenge of data silos, where valuable data is fragmented across various organizations or devices and cannot be easily combined due to regulatory, competitive, or logistical barriers. By enabling collaborative model training on these disparate datasets, FL unlocks new opportunities for knowledge discovery and model improvement that would otherwise be unattainable. Moreover, FL leverages the computational resources at the edge, reducing the need for extensive cloud infrastructure and minimizing communication bandwidth, especially when dealing with large datasets. This distributed nature also enhances system robustness, as the failure of a single client does not cripple the entire training process[5].

Federated learning has rapidly found diverse applications across a multitude of sectors. In healthcare, it facilitates the development of advanced diagnostic models by allowing hospitals to collaboratively train on patient data without compromising individual privacy, leading to more accurate disease detection and personalized treatment plans[3]. The financial industry utilizes FL for fraud detection and risk assessment, enabling banks to share insights from their transaction data while maintaining customer confidentiality[2]. In the realm of recommendation systems, platforms can offer highly personalized content suggestions by learning from user interactions directly on devices, without centralizing sensitive user behavior data[4]. Furthermore, FL is pivotal in advancing smart city initiatives, autonomous vehicles, and the Internet of Things (IoT), where it enables intelligent decision-making at the edge, optimizing resource allocation and enhancing operational efficiency[6, 27]. This paper aims to provide a comprehensive review of federated learning, delving into its foundational concepts, evolutionary trajectory, the critical challenges it currently faces, and its promising future directions. By offering an in-depth analysis, this review seeks to equip researchers and practitioners with a nuanced understanding of FL's principles and its potential to reshape the landscape of privacy-preserving artificial intelligence.

2. FEDERATED LEARNING OVERVIEW

At its core, federated learning operates on a collaborative yet decentralized principle, fundamentally altering the traditional machine learning paradigm. The process is typically orchestrated by a central coordinating server, which initiates the learning cycle by distributing an initial or current version of a global model to a multitude of participating client devices. These clients, which can range from mobile phones and wearable devices to institutional servers and IoT sensors, then undertake the crucial task of local model training. Each client leverages its proprietary, local dataset—data that never leaves the device—to refine the received model. This local training phase involves computing model updates, such as gradients or updated model parameters, based on the client's unique data distribution.

Upon completion of local training, instead of transmitting their raw data, clients securely send only these computed model updates back to the central server. The server then performs an aggregation step, combining the updates received from all participating clients to produce a refined global model. This aggregation process is designed to synthesize the collective knowledge gained from the distributed datasets while preserving the privacy of individual data points. Once the global model is updated, it is redistributed to the clients for the next round of local training, and this iterative cycle continues until the model converges to a satisfactory performance level or a predefined number of communication rounds are completed. This iterative exchange of model updates, rather than raw data, is the cornerstone of federated learning's privacy-preserving capabilities.

Key Distinctions from Traditional Distributed Learning

While federated learning is a form of distributed machine learning, it possesses several critical distinctions that set it apart from conventional distributed training approaches:

1. Emphasis on Privacy Preservation : The most salient difference lies in the paramount importance placed on privacy. In federated learning, client devices retain absolute control and ownership over their private data. The central server, acting solely as an orchestrator, neither collects nor stores any raw client data. This contrasts sharply with traditional distributed machine learning, where a central node or cluster typically manages and has full access to all partitioned data across the distributed system. In such traditional setups, data is often sharded and distributed to worker nodes, but the central authority still maintains a comprehensive view and control over the entire dataset.
2. Heterogeneity and Inclusivity of Client Devices : Federated learning is designed to accommodate a wide spectrum of client devices, each potentially possessing varying computational capabilities,

storage capacities, network bandwidths, and data volumes. This high degree of inclusivity means that participants can include resource-constrained mobile devices, smart home appliances, industrial sensors, or even diverse organizational servers. Traditional distributed machine learning environments, conversely, are typically deployed in more homogeneous and controlled settings, such as data centers or high-performance computing clusters. In these environments, worker nodes are generally uniform in their computational power and resources, ensuring predictable performance and easier management.

3. Addressing Distinct Challenges : Federated learning extends the foundational framework of distributed systems to tackle challenges primarily related to data privacy, data silos, and efficient utilization of edge computing resources. Its focus is on enabling collaborative model training when data cannot be centralized due to privacy concerns, regulatory restrictions, or logistical complexities. Traditional distributed machine learning, on the other hand, primarily aims to enhance computational efficiency and scalability in big data scenarios. Its objective is to accelerate model training and reduce time costs by parallelizing tasks and distributing data across multiple nodes, assuming data can be freely moved and aggregated.

4. Data Distribution Characteristics : In an ideal traditional distributed learning setting, data is often assumed to be independently and identically distributed (IID) across all nodes, or at least carefully partitioned to approximate IID conditions. Federated learning, however, inherently deals with non-IID data distributions. Each client's local dataset reflects its unique usage patterns, demographics, or environmental factors, leading to statistical heterogeneity across clients. This non-IID nature poses significant algorithmic challenges for model convergence and generalization, which FL algorithms must explicitly address.

These distinctions highlight federated learning's unique position as a privacy-preserving, distributed machine learning paradigm tailored for real-world scenarios where data is decentralized, heterogeneous, and sensitive. Its architectural flexibility and inherent privacy features make it a compelling solution for a growing number of applications across diverse industries.

3. DEVELOPMENT OF FEDERATED LEARNING

The genesis of federated learning can be traced back to 2016, when researchers at Google first introduced the concept [1]. Their seminal work laid the groundwork for a novel approach to machine learning that allowed models to be trained on decentralized client data without the necessity of transmitting raw data to a central server, thereby safeguarding user privacy [24]. The Federated Averaging (FedAvg) algorithm, proposed in this foundational paper, quickly became the most widely adopted method in federated learning. In FedAvg, the central server's role is simplified to merely aggregating the model parameters (e.g., weights and biases of a neural network) uploaded by participating client devices, typically by computing their weighted average. This design elegantly bypasses the need for the central server to engage in direct model training or data management, significantly enhancing privacy.

However, the real-world deployment of federated learning soon revealed a critical challenge: the non-independent and identically distributed (non-IID) nature of client data. Unlike controlled laboratory settings where data can often be assumed to be IID, data generated by diverse client devices in heterogeneous environments is inherently non-IID. This statistical heterogeneity can lead to significant performance degradation and unstable model convergence in vanilla FedAvg. In response, a wave of research has focused on developing more robust and efficient aggregation strategies and local optimization techniques to mitigate the adverse effects of non-IID data. For instance, FedProx [7] introduced a proximal term to the local objective function, penalizing deviations between the local model and the global model. This regularization helps to stabilize training and improve convergence in non-IID settings by encouraging local models to stay closer to the global consensus. Similarly, SCAFFOLD [8] proposed a novel control variate approach to correct for client drift caused by local data heterogeneity, aiming to ensure that local updates are more aligned with the global objective. FedDyn [9] further advanced this by incorporating dynamic regularization based on historical model updates, providing a more adaptive mechanism to manage the bias introduced by non-IID data, rather than relying solely on the current model state.

Another significant evolutionary step in federated learning is the emergence of personalized federated learning [10]. Recognizing that a single global model might not optimally serve all diverse clients, personalized FL aims to tailor models to individual client needs while still benefiting from collaborative learning. This approach seeks a balance between global generalization and local specialization. Various strategies have been explored to achieve personalization. For example, LG-FedAvg [11] proposed a method where the top layers of a model are treated as shared parameters, while the bottom layers are personalized, allowing for both global knowledge transfer and local adaptation. FedRod [12] introduced the concept of maintaining a private personalized classifier on each client in addition to sharing the entire private model, enabling more nuanced personalization. FedBABU [13] explored a phased approach, where clients continuously update

and share the bottom-layer parameters of their private models in the initial stages, followed by fine-tuning the top layers to acquire personalized models later. Personalized federated learning represents a crucial advancement, as it not only facilitates the training of models highly adapted to individual data distributions but also maintains the integrity and benefits of global federated optimization.

The rapid advancements in deep learning and the advent of large-scale models have also spurred interest in federated learning for large models [14]. Training massive models, such as large language models or vision transformers, typically requires immense computational resources and vast datasets, often centralized. Federated learning offers a compelling alternative by enabling the collaborative training of these large models across distributed edge devices, potentially leveraging their collective data without centralizing it. This area of research is still nascent but holds immense promise for democratizing access to powerful AI models and enabling their deployment in privacy-sensitive environments. Furthermore, the integration of federated learning with blockchain technology [15] has gained traction. Blockchain can provide a decentralized, immutable, and transparent ledger for recording model updates and client contributions, thereby enhancing the security, trustworthiness, and traceability of federated learning processes. This synergy can further bolster privacy protection and provide verifiable audit trails, addressing concerns about data integrity and malicious participants in FL ecosystems.

4. CURRENT CHALLENGES IN FEDERATED LEARNING

Despite its significant advantages and rapid advancements, federated learning is not without its inherent challenges. These obstacles often stem from the decentralized nature of the paradigm and the complexities of real-world data distributions and network environments. Addressing these challenges is crucial for the widespread adoption and robust performance of federated learning systems.

4.1. Data Heterogeneity (Non-IID Data)

One of the most pervasive and challenging issues in federated learning is data heterogeneity, often referred to as the non-independent and identically distributed (non-IID) nature of client data. In an idealized federated learning scenario, where data across all participating clients is IID, classical federated learning algorithms like FedAvg can achieve excellent model performance and rapid convergence. However, in practical applications, client data is rarely IID. Each client's local dataset is typically generated from its unique environment, user behavior, or demographic characteristics, leading to significant statistical differences in data distributions across clients. This non-IID characteristic manifests in several ways:

· Feature Distribution Skew : Different clients may have data with varying feature distributions. For example, in a medical imaging task, one hospital might have a higher prevalence of a certain disease compared to another.

· Label Distribution Skew : Clients might have different distributions of labels. A mobile phone user might primarily interact with certain applications, leading to a skewed distribution of app usage data.

· Quantity Skew : The amount of data available on each client can vary significantly, with some clients possessing vast datasets and others having very limited data.

· Concept Drift : The underlying data distribution on a client might change over time, leading to a dynamic non-IID scenario.

This data heterogeneity poses a severe challenge to model convergence and generalization. When clients train on vastly different local data distributions, their local model updates can pull the global model in conflicting directions, leading to slow convergence, oscillations, or even divergence. The aggregated global model may struggle to perform well across all clients, particularly on those with minority data distributions. To counteract these issues, researchers have explored various strategies. Customizing personalized parameters, as seen in personalized federated learning approaches, aims to allow each client to adapt the global model to its local data characteristics. Another promising direction is knowledge distillation, where a global model distills knowledge to local models or vice versa, enabling efficient transfer of information while respecting data privacy. However, both personalized models and knowledge distillation often introduce additional computational overhead, requiring more complex algorithms and potentially longer training times, which can be a significant concern for resource-constrained edge devices.

4.2. Straggler Effect and Client Selection

The assumption of global participation, where all clients contribute to every round of federated learning, is often unrealistic in real-world deployments. The straggler effect, a prominent challenge in federated learning, arises from the inherent heterogeneity of client devices in terms of hardware capabilities, network bandwidth, and data volume. These disparities can significantly impact the efficiency and convergence of the federated training process.

Specifically:

· Hardware Differences : Clients possess diverse computational powers, ranging from high-end servers in cross-silo FL to low-power mobile devices in cross-device FL. This leads to varying local training speeds, with slower devices becoming bottlenecks.

· Network Bandwidth and Latency : The efficiency of model download from and upload to the central

server is heavily dependent on the client's network connectivity. Clients with poor or intermittent network connections can delay the aggregation process.

· Data Volume Differences : Clients with larger datasets require more computational resources and time for local model updates compared to those with smaller datasets. This can lead to inconsistent convergence rates among private models.

These less efficient client devices are termed 'stragglers.' Their delayed participation or failure to complete local training within a given timeframe can severely disrupt the model aggregation process at the central server, impacting both the speed and quality of the global model. If the central server waits for all clients, the overall training time can be significantly prolonged, negating the benefits of distributed computation. If it proceeds without stragglers, the aggregated model might be biased or less representative of the overall data distribution.

To address the straggler effect, various strategies have been proposed. Asynchronous update strategies, where the central server does not wait for all clients to complete their local training before aggregation, can mitigate delays. However, purely asynchronous approaches can lead to issues like model staleness, where updates from slower clients arrive too late to be fully relevant to the current global model state. Other approaches involve sophisticated client selection mechanisms, where the central server strategically chooses a subset of clients for each training round based on factors like their computational resources, network conditions, data quality, or even their historical reliability. While these methods can improve efficiency, they introduce complexity and may not always maximize the overall benefit, potentially leading to biases if certain client data distributions are consistently underrepresented.

4.3. Privacy Protection and Security

While federated learning is inherently designed with privacy in mind, it is not impervious to privacy breaches or security threats. The very act of sharing model updates, even without raw data, can inadvertently leak sensitive information [26]. The primary mechanisms for privacy protection in federated learning include:

1. Inherent Design Advantage : The foundational principle of FL—that raw data never leaves the client device—is its first and most significant privacy safeguard. The central server only receives aggregated model updates, not individual data points.
2. Differential Privacy (DP) : Differential privacy is a rigorous mathematical framework that provides strong privacy guarantees by introducing carefully calibrated noise into the model updates before they are sent to the central server [18]. This noise makes it statistically difficult for an adversary to infer information about any single individual's data from the aggregated updates. While highly effective, implementing differential privacy often comes with a trade-off: the added noise can reduce the accuracy of the trained model, and determining the optimal level of noise is a critical challenge.
3. Homomorphic Encryption (HE) : Homomorphic encryption allows computations to be performed on encrypted data without decrypting it [19]. In federated learning, this means that clients can encrypt their model updates before sending them to the server, and the server can aggregate these encrypted updates without ever seeing the unencrypted values. Only the final aggregated model, or specific results, are decrypted. Homomorphic encryption offers a very high level of privacy, but its main drawback is the significant computational and communication overhead it introduces, making it resource-intensive for many practical FL deployments, especially on edge devices.

Beyond these primary privacy-enhancing technologies, federated learning systems are also vulnerable to various security threats, including:

· Model Poisoning Attacks : Malicious clients can intentionally send corrupted or adversarial model updates to the central server, aiming to degrade the global model's performance or introduce backdoors.

· Data Poisoning Attacks : Although raw data is not shared, an attacker might inject malicious data into their local dataset to influence the training process.

· Inference Attacks : Even with privacy mechanisms, sophisticated adversaries might attempt to infer sensitive information about individual clients or their data by analyzing the shared model updates or the global model itself. This includes membership inference attacks (determining if a specific data point was part of the training set) and property inference attacks (inferring properties of the training data).

· Sybil Attacks : An attacker might create multiple fake client identities to gain disproportionate influence over the global model.

Addressing these privacy and security challenges requires a multi-faceted approach, often combining cryptographic techniques, differential privacy, secure multi-party computation (SMC), and robust aggregation algorithms. The ongoing research in this area focuses on developing more efficient and lightweight privacy-preserving strategies that can be deployed on resource-constrained client devices without significantly compromising model utility or incurring excessive computational and communication costs. The balance between privacy, utility, and efficiency remains a central research problem in federated learning.

5. FUTURE OUTLOOK OF FEDERATED LEARNING

Federated learning is a rapidly evolving field with immense potential to reshape how machine learning models are developed and deployed, particularly in privacy-sensitive and data-rich environments. As the technology matures, several key areas are poised for significant advancements and research focus.

5.1. Personalized Federated Learning

The concept of personalized federated learning has already demonstrated considerable effectiveness in bridging the gap between a single global model and the diverse needs of individual clients. While current research often assumes a homogeneous model structure across all global client devices, the future of personalized FL lies in pushing this adaptability further. This involves developing strategies that can dynamically and adaptively match appropriate model architectures and learning paradigms to the unique conditions and data characteristics of each client device. A critical challenge in this pursuit is designing mechanisms for the central server to effectively integrate updates from heterogeneous models, where clients might be training different model types or architectures. This could involve meta-learning approaches, multi-task learning, or advanced knowledge transfer techniques that can distill insights from diverse local models into a coherent global representation.

Furthermore, the development of active adjustment strategies for client devices is a promising avenue. Instead of passively receiving global model updates, clients could autonomously adjust their local hyperparameters, learning rates, or even model architectures based on their historical training performance, data drift, or specific task requirements. This would empower clients to optimize their local learning processes more effectively, leading to faster convergence, improved local model performance, and better overall resource utilization within the federated ecosystem. Research into reinforcement learning or adaptive control mechanisms for client-side optimization could play a pivotal role in realizing this vision.

5.2. Federated Learning and Large Models

The recent explosion in the scale and capabilities of large models, such as large language models (LLMs) and foundation models, has ignited significant interest in integrating them within the federated learning framework. Superficially, large models and federated learning appear to have conflicting philosophies: FL advocates for lightweight models to minimize computational, storage, and communication overhead on edge devices, whereas large models inherently rely on massive architectures and billions of parameters to process and understand high-dimensional data. However, the synergy between these two fields holds transformative potential.

One promising direction involves using federated learning for information integration, where a central large model acts as a powerful aggregator and knowledge refiner. In this scenario, edge devices could perform initial data processing or train smaller, specialized models locally. The insights or distilled knowledge from these local models would then be transmitted to a central large model, which would perform high-level information extraction, learning, and generalization. Subsequently, this central large model could generate more efficient, lightweight models through techniques like knowledge distillation, which are then deployed back to the edge devices. This approach leverages the strengths of both: the privacy-preserving and distributed nature of FL for data access, and the powerful generalization capabilities of large models for complex pattern recognition and knowledge synthesis.

Another critical area of research is enabling the training of large models directly on resource-constrained client devices within a federated setting. This necessitates significant advancements in model compression techniques, including pruning, quantization, and distillation. By drastically reducing the size and computational footprint of large models, it becomes feasible to train them on edge devices. Federated learning would then facilitate the collaborative aggregation of updates from these compressed local models, enabling the collective intelligence of distributed data to contribute to the development of powerful, yet deployable, large models. This could unlock unprecedented opportunities for on-device AI, personalized large language models, and efficient deployment of advanced AI capabilities in privacy-sensitive edge environments.

6. DISCUSSION

Federated learning, while offering a compelling solution to privacy concerns and data silo challenges, is still a nascent field with numerous avenues for deeper exploration and refinement. The discussions surrounding its practical deployment often revolve around the delicate balance between privacy, model utility, communication efficiency, and computational feasibility across heterogeneous client environments. The inherent non-IID nature of data in real-world FL scenarios remains a central point of contention and active research. While personalized FL approaches and advanced aggregation techniques have shown promise in mitigating the negative impacts of data heterogeneity, the optimal strategies often depend on the specific application domain and the degree of data divergence among clients. Further research is needed to develop adaptive algorithms that can

dynamically adjust to varying levels of non-IIDness and provide robust performance guarantees.

The straggler effect, stemming from the diverse computational and network capabilities of participating devices, poses a significant hurdle to the efficiency and scalability of FL systems. While asynchronous update mechanisms and intelligent client selection strategies offer partial solutions, they often introduce new complexities, such as model staleness or potential biases in client representation. Future discussions will likely focus on more sophisticated resource management techniques, perhaps incorporating predictive models to anticipate and mitigate straggler behavior, or developing incentive mechanisms to encourage consistent participation from all clients. The trade-offs between system responsiveness and model convergence in the presence of stragglers will continue to be a critical area of investigation.

Privacy and security, though foundational to FL, are not fully resolved challenges. The vulnerability of FL systems to various attacks, including model poisoning, data inference, and Sybil attacks, necessitates continuous innovation in defense mechanisms. While differential privacy and homomorphic encryption provide strong theoretical guarantees, their practical implementation often comes with significant computational overhead or a reduction in model accuracy. The discussion needs to shift towards developing more lightweight, efficient, and composable privacy-preserving techniques that can be seamlessly integrated into diverse FL architectures without compromising utility. Furthermore, the development of robust auditing and verification mechanisms to ensure the integrity and trustworthiness of aggregated models will be paramount for building confidence in FL systems, especially in highly regulated industries.

Beyond these technical challenges, the broader implications of federated learning on data governance, regulatory frameworks, and ethical considerations warrant extensive discussion. As FL becomes more prevalent, questions regarding data ownership, accountability for model biases, and the potential for misuse of aggregated intelligence will become increasingly important. Establishing clear legal and ethical guidelines for the deployment of FL systems will be crucial for fostering public trust and ensuring responsible innovation. The interdisciplinary nature of these challenges underscores the need for collaboration among machine learning researchers, cryptographers, legal experts, and policymakers to collectively shape the future of privacy-preserving AI.

6.1. Ethical Considerations and Regulatory Landscape

Beyond the technical intricacies, the widespread adoption of federated learning introduces a complex array of ethical considerations and necessitates a robust regulatory framework. While FL inherently addresses privacy by keeping raw data localized, it does not automatically resolve all ethical dilemmas. For instance, questions arise regarding algorithmic fairness and bias. If the training data on participating clients is inherently biased, the aggregated global model can perpetuate and even amplify these biases, leading to discriminatory outcomes, particularly in sensitive applications like healthcare or finance. Ensuring fairness across diverse client populations, especially when data distributions are non-IID, is a critical ethical challenge that requires proactive measures, such as fairness-aware aggregation algorithms and rigorous auditing mechanisms [20].

Another ethical concern revolves around accountability. In a decentralized training paradigm, pinpointing responsibility for model errors, biases, or privacy breaches becomes significantly more complex. Who is accountable when a federated model makes a harmful decision: the central orchestrator, the contributing clients, or a combination thereof? Clear guidelines and legal frameworks are needed to delineate responsibilities and establish mechanisms for redress. Furthermore, the potential for malicious actors to inject poisoned data or model updates, even with privacy-preserving techniques, raises questions about the trustworthiness of the aggregated model and the need for robust verification processes [21].

The evolving regulatory landscape, driven by privacy-centric legislations like GDPR in Europe and CCPA in California, significantly influences the development and deployment of FL. Federated learning is often seen as a promising tool for compliance with these regulations, as it minimizes data transfer and central storage. However, the nuances of FL, such as the potential for inference attacks or the aggregation of sensitive model updates, mean that mere adoption of FL does not guarantee full compliance [25]. Regulators and policymakers are increasingly grappling with how to adapt existing data protection laws to the unique characteristics of FL, particularly concerning data ownership, consent mechanisms for model training, and the right to be forgotten in a distributed learning context. The development of standardized protocols and best practices for FL deployment, alongside clear legal interpretations, will be crucial for fostering trust and accelerating its responsible integration into various industries [22].

6.2. Interoperability and Standardization

The current federated learning ecosystem is characterized by a proliferation of diverse frameworks, algorithms, and deployment strategies, leading to significant challenges in interoperability and standardization. Different research groups and companies often develop their own proprietary or open-source FL platforms, each with unique APIs, data formats, and communication protocols. This fragmentation hinders the seamless integration of

FL solutions across different organizations and limits the ability to benchmark and compare the performance of various FL algorithms effectively. The lack of universal standards makes it difficult for new entrants to adopt FL, increases development costs, and impedes the creation of a truly collaborative and scalable FL ecosystem.

Standardization efforts are crucial to address these issues. This includes developing common data exchange formats, standardized communication protocols for model updates, and unified APIs for interacting with FL platforms. Such standards would facilitate the creation of modular and interoperable FL components, allowing researchers and practitioners to easily combine different algorithms, privacy-preserving techniques, and hardware configurations. Furthermore, the establishment of standardized benchmarking datasets and evaluation metrics, particularly for non-IID scenarios and adversarial attacks, is essential for objectively assessing the performance and robustness of FL systems. Collaborative initiatives involving academia, industry, and regulatory bodies are necessary to drive these standardization efforts, ensuring that federated learning can evolve into a mature and widely adopted technology with a robust and interconnected ecosystem [23].

7. OPEN ISSUES AND FUTURE RESEARCH DIRECTIONS

Despite the significant progress in federated learning, several open issues and promising research directions remain that warrant further investigation to unlock its full potential and address its limitations. These areas represent fertile ground for future innovation and will be critical for the widespread adoption of FL in diverse real-world applications.

7.1. Robustness to Data Heterogeneity and Non-IID Data

While personalized federated learning and various regularization techniques have been proposed to mitigate the effects of non-IID data, a universally robust solution remains elusive. Future research should focus on:

- **Adaptive Personalization Strategies**: Developing more sophisticated adaptive personalization methods that can dynamically adjust the degree of personalization based on the client's data characteristics, computational resources, and the specific task at hand. This could involve meta-learning for personalization or reinforcement learning to guide personalized model updates.
- **Fairness in Non-IID Settings**: Ensuring fairness across clients, especially when data distributions are highly skewed. Non-IID data can lead to models that perform exceptionally well on data from dominant clients but poorly on data from minority clients. Research is needed to develop fairness-aware FL algorithms that can guarantee equitable performance across all participants.
- **Theoretical Understanding of Non-IID Effects**: Deepening the theoretical understanding of how non-IID data impacts convergence, generalization, and privacy in FL. This includes developing tighter theoretical bounds and more accurate predictive models for FL performance under various non-IID conditions.

7.2. Communication Efficiency and Scalability

Communication overhead remains a major bottleneck, especially in cross-device FL with a large number of resource-constrained clients. Future research should explore:

- **Advanced Compression Techniques**: Developing more aggressive yet lossless or near-lossless model update compression techniques, including quantization, sparsification, and knowledge distillation, to reduce the amount of data transmitted between clients and the server.
- **Asynchronous and Semi-Asynchronous FL**: Further optimizing asynchronous and semi-asynchronous FL algorithms to handle stragglers more effectively without compromising model quality or introducing significant staleness. This could involve dynamic weighting of client contributions based on their update freshness.
- **Hierarchical Federated Learning**: Investigating hierarchical FL architectures, where multiple layers of aggregation are introduced (e.g., local aggregators within a region before sending to a central server). This can reduce the load on the central server and improve communication efficiency in large-scale deployments.

7.3. Enhanced Privacy and Security Mechanisms

Despite the inherent privacy benefits, FL systems are still susceptible to various attacks. Future research needs to focus on:

- **Lightweight Cryptographic Primitives**: Developing more efficient and lightweight cryptographic techniques, such as secure multi-party computation (SMC) and homomorphic encryption (HE), that can be practically deployed on edge devices without prohibitive computational or communication costs.
- **Robustness against Adversarial Attacks**: Designing FL systems that are inherently more robust against various adversarial

attacks, including model poisoning, data poisoning, and inference attacks. This involves developing robust aggregation rules, anomaly detection mechanisms, and secure client authentication protocols.

- **Auditing and Explainability**: Enhancing the transparency and explainability of FL models, particularly in sensitive applications like healthcare and finance. This includes developing methods to audit the contributions of individual clients and to explain model decisions in a privacy-preserving manner.

7.4. Integration with Emerging Technologies

Federated learning's potential can be further amplified by its integration with other cutting-edge technologies:

- **FL and Edge AI**: Deepening the integration of FL with edge computing paradigms to enable more intelligent and autonomous decision-making at the network edge. This includes optimizing FL algorithms for deployment on specialized edge hardware and developing frameworks for seamless FL deployment on edge devices.
- **FL and Blockchain**: Further exploring the synergy between FL and blockchain for enhanced security, transparency, and incentive mechanisms. Blockchain can provide a decentralized and immutable ledger for FL operations, facilitating trust and accountability among participants.
- **FL and Large Language Models (LLMs)**: Addressing the unique challenges of training and deploying LLMs in a federated setting. This includes developing efficient methods for federated fine-tuning of LLMs, managing the massive model sizes, and ensuring privacy during the training of such powerful models.

7.5. Real-world Deployment and Standardization

Moving beyond theoretical advancements, practical deployment and standardization are crucial for FL's widespread adoption:

- **Benchmarking and Evaluation**: Establishing standardized benchmarks and evaluation metrics for FL systems that accurately reflect real-world conditions, including non-IID data, heterogeneous clients, and various attack scenarios.
- **Framework Development**: Continuing the development of user-friendly and robust open-source FL frameworks that abstract away much of the underlying complexity, making FL more accessible to researchers and practitioners.
- **Regulatory and Ethical Guidelines**: Collaborating with policymakers and ethicists to develop clear regulatory frameworks and ethical guidelines for FL deployment, ensuring responsible innovation and addressing societal concerns related to data privacy and algorithmic bias.

By addressing these open issues and pursuing these research directions, federated learning can evolve into an even more powerful and pervasive technology, driving the next generation of privacy-preserving and collaborative artificial intelligence systems.

8. CONCLUSION

Federated learning has firmly established itself as a pivotal paradigm in the evolution of artificial intelligence, offering a compelling response to the dual challenges of data privacy and data fragmentation. Since its introduction, FL has not only garnered significant academic interest but has also seen practical applications across a diverse range of industries, including healthcare, finance, and telecommunications. Its ability to facilitate collaborative model training on decentralized datasets without compromising user privacy has unlocked new frontiers for AI innovation, enabling the development of more robust and personalized models. This review has provided a comprehensive overview of the federated learning landscape, from its foundational concepts and evolutionary trajectory to the critical challenges that continue to shape its development. We have explored the various architectural and algorithmic nuances of FL, including the critical distinctions from traditional distributed learning, the ongoing efforts to address data heterogeneity and the straggler effect, and the multifaceted approaches to bolstering privacy and security. However, the journey towards seamless and widespread adoption of federated learning is far from over. The open issues and research directions highlighted in this paper underscore the complexity and dynamism of the field. The challenges of non-IID data, communication efficiency, and robust security are not merely technical hurdles but fundamental research questions that require continued and concerted efforts from the global research community. The future of federated learning will likely be characterized by a move towards more adaptive, personalized, and resource-aware systems that can intelligently navigate the complexities of real-world deployments. The integration of FL with emerging technologies such as edge AI, blockchain, and large language models will further expand its capabilities and application domains, paving the way for a new generation of intelligent, decentralized, and privacy-preserving systems. Federated learning represents a significant

step forward in our quest to build more responsible and effective AI. By embracing a decentralized and collaborative approach, FL not only addresses the pressing need for data privacy but also democratizes access to advanced machine learning capabilities. As the field continues to mature, the ongoing dialogue between researchers, practitioners, policymakers, and the public will be crucial in shaping a future where the immense potential of federated learning is realized in a manner that is both ethically sound and technologically robust. The continued exploration of the open issues discussed in this review will be instrumental in driving this evolution and ensuring that federated learning remains a cornerstone of privacy-preserving artificial intelligence for years to come.

REFERENCES

[1] McMahan, B., Moore, E., Ramage, D., et al. (2017) Communication-Efficient Learning of Deep Networks from Decentralized Data. arXiv: 1602.05629.

[2] Byrd, D. and Polychroniadou, A. (2020) Differentially Private Secure Multi-Party Computation for Federated Learning in Financial Applications. Proceedings of the First ACM International Conference on AI in Finance, New York, 15-16 October 2020, 1-9. https://doi.org/10.1145/3383455.3422562

[3] Xu, J., Glicksberg, B.S., Su, C., Walker, P., Bian, J. and Wang, F. (2020) Federated Learning for Healthcare Informatics. Journal of Healthcare Informatics Research, 5, 1-19. https://doi.org/10.1007/s41666-020-00082-4

[4] Muhammad, K., Wang, Q., O'Reilly-Morgan, D., Tragos, E., Smyth, B., Hurley, N., et al. (2020) FedFast: Going Beyond Average for Faster Training of Federated Recommender Systems. Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 6-10 July 2020, 1234-1242. https://doi.org/10.1145/3394486.3403176

[5] Nguyen, D.C., Ding, M., Pathirana, P.N., Seneviratne, A., Li, J. and Vincent Poor, H. (2021) Federated Learning for Internet of Things: A Comprehensive Survey. IEEE Communications Surveys & Tutorials, 23, 1622-1658. https://doi.org/10.1109/comst.2021.3075439

[6] Zeng, T., Semiari, O., Chen, M., Saad, W. and Bennis, M. (2022) Federated Learning on the Road Autonomous Controller Design for Connected and Autonomous Vehicles. IEEE Transactions on Wireless Communications, 21, 10407-10423. https://doi.org/10.1109/twc.2022.3183996

[7] Li, T., Sahu, A.K., Zaheer, M., et al. (2020) Federated Optimization in Heterogeneous Networks. Machine Learning and Systems (MLSys), 2, 429-450.

[8] Karimi Reddy, S.P., Kale, S., Mohri, M., et al. (2020) SCAFFOLD: Stochastic Controlled Averaging for Federated Learning. arXiv: 1910.06378.

[9] Acar D A E, Zhao Y, Navarro R M, et al. (2021) Federated Learning Based on Dynamic Regularization. arXiv: 2111.04263.

[10] Tan, A.Z., Yu, H., Cui, L. and Yang, Q. (2023) Towards Personalized Federated Learning. IEEE Transactions on Neural Networks and Learning Systems, 34, 9587-9603. https://doi.org/10.1109/tnnls.2022.3160699

[11] Liang, P.P., Liu, T., Ziyin, L., et al. (2019) Think Locally, Act Globally: Federated Learning with Local and Global Representations. https://arxiv.org/abs/2001.01523

[12] Chen, H.Y. and Chao, W.L. (2022) On Bridging Generic and Personalized Federated Learning for Image Classification. arXiv: 2107.00778.

[13] Oh, J., Kim, S. and Yun, S.Y. (2022) FedBABU: Towards Enhanced Representation for Federated Image Classification. arXiv: 2106.06042.

[14] Chen, C., Feng, X., Zhou, J., Yin, J. and Zheng, X. (2023) Federated Large Language Model: A Position Paper. arXiv: 2307.08925.

[15] Wang, Z. and Hu, Q. (2021) Blockchain-Based Federated Learning: A Comprehensive Survey. arXiv: 2110.02182.

[16] Li, Q., Diao, Y., Chen, Q. and He, B. (2022) Federated Learning on Non-IID Data Silos: An Experimental Study. 2022 IEEE 38th International Conference on Data Engineering (ICDE), Kuala Lumpur, 9-12 May 2022, 965-978. https://doi.org/10.1109/icde53745.2022.00077

[17] Chai, Z., Chen, Y., Zhao, L., Cheng, Y. and Rangwala, H. (2020) FedAT: A Communication-Efficient Federated Learning Method with Asynchronous Tiers under Non-IID Data.

[18] Wei, K., Li, J., Ding, M., Ma, C., Yang, H.H., Farokhi, F., et al. (2020) Federated Learning with Differential Privacy: Algorithms and Performance Analysis. IEEE Transactions on Information Forensics and Security, 15, 3454-3469. https://doi.org/10.1109/tifs.2020.2988575

[19] Wibawa, F., Catak, F.O., Kuzlu, M., Sarp, S. and Cali, U. (2022) Homomorphic Encryption and Federated Learning Based Privacy-Preserving CNN Training: COVID-19 Detection Use-Case. EICC 2022: Proccedings of the European Interdisciplinary Cybersecurity Conference, Barcelona, 15-16 June 2022, 85-90. https://doi.org/10.1145/3528580.3532845

[20] Li, L., Fan, Y., Feng, M., et al. (2020) A Survey on Federated Learning: Concepts, Applications and Future Directions. IEEE Access, 8, 120360-120377. https://doi.org/10.1109/ACCESS.2020.3007103

[21] Kairouz, P., McMahan, H.B., Avent, B., et al. (2021) Advances and Open Problems in Federated

Learning. Foundations and Trends® in Machine Learning, 14, 1-210. https://doi.org/10.1561/2200000083

[22] Warnat-Herresthal, S., Schultze, H., Shastry, K.L., et al. (2021) Swarm Learning for Decentralized and Confidential Clinical Machine Learning. Nature Medicine, 27, 1085-1093. https://doi.org/10.1038/s41591-021-01385-y

[23] Mothukuri, V., Parizi, R.M., Pouriyeh, S., et al. (2021) A Survey on Federated Learning: Challenges, Methods, and Future Directions. Computer Networks, 204, 108698. https://doi.org/10.1016/j.comnet.2021.108698

[24] Wajahat, A., He, J., Zhu, N., Mahmood, T., Saba, T., Khan, A.R. and Alamri, F.S., 2024. Outsmarting Android Malware with Cutting-Edge Feature Engineering and Machine Learning Techniques. *Computers, Materials & Continua*, *79*(1). https:// 10.32604/cmc.2024.047530

[25] Qureshi, S., Li, J., Akhtar, F., Tunio, S., Khand, Z.H. and Wajahat, A., 2021. Analysis of challenges in modern network forensic framework. *Security and Communication Networks*, *2021*(1), p.8871230. https://doi.org/10.1155/2021/8871230

[26] Wajahat, A., He, J., Zhu, N., Mahmood, T., Nazir, A., Pathan, M.S., Qureshi, S. and Ullah, F., 2023. An adaptive semi-supervised deep learning-based framework for the detection of Android malware. *Journal of Intelligent & Fuzzy Systems*, *45*(3), pp.5141-5157. https://doi.org/10.3233/JIFS-231969

[27] Alalloush, H., & Ali, W. (2023). API Malware Analysis: Exploring Detection And Forensics Strategies For Secure Software Development. Journal of Intelligent Systems and Applied Data Science, 1(1).

Journal of Intelligent System and Applied Data Science (JISADS)

Journal homepage : https://www.jisads.com
ISSN (2974-9840) Online

DEEP LEARNING FOR OFFLINE SIGNATURE VERIFICATION: A NOVEL MULTI-CHANNEL FEATURE FUSION NETWORK

*Harshal Hemane[1] *, Anuradha Kasangottuwar[2]*

*[1] * E&TC Engineering Department, DACOE, Karad, India*
[2]PES Modern COE in department of E&TC Engineering, Pune, Maharashtra, India

ABSTRACT

This paper addresses the critical challenge of offline signature verification, a task crucial for authenticating documents and identities. Existing deep learning approaches, primarily deep metric learning with Siamese networks and two-channel discriminative methods, face limitations. While Siamese networks excel at feature extraction, their reliance on Euclidean distance can overlook subtle directional and scaling information, hindering the capture of intricate feature relationships. Conversely, two-channel discriminative methods, though effective in initial dissimilarity assessment, often suffer from significant feature loss due to early image fusion. To overcome these challenges, we propose the Multi-channel Feature Fusion Network , a novel writer-independent model for handwritten signature verification. The proposed framework leverages a quadruple Siamese network and a dual inverse discriminative attention mechanism for robust feature extraction and enhancement from both original and inverse grayscale images. These rich, multi-dimensional features are then integrated through an innovative channel fusion process. Finally, an ACMix-based discriminative module is employed to determine image similarity with high precision. Comprehensive experiments on four diverse language signature demonstrate the superior efficacy and promising potential of the framework, affirming its advantages over current methodologies.

Keywords: Offline handwritten signature verification, deep learning, channel fusion

1. INTRODUCTION

In contemporary society, signature handwriting verification, as one of the crucial forensic methods, is widely applied in various fields such as law, insurance, and culture [15,20,10,19,41]. Due to the uniqueness, stability, and reliability of signature handwriting, it serves as an important basis for authenticating documents and confirming identities. However, with the continuous advancement of technology, signature handwriting examination also faces numerous challenges. The origin of signature handwriting can be traced back to ancient times when people used various symbols and graphics to sign. With the development of paper and ink, people began to use handwritten signatures. As early as 439 AD, the Roman Empire used signatures to verify the authenticity of documents. However, it was not until the early 20th century that signature handwriting began to attract research attention. During this period, disciplines such as psychology and statistics began to be applied to the study of signature handwriting, providing a theoretical basis for signature handwriting examination.

Signature handwriting plays an important role in multiple fields. In the legal field, signature handwriting is a crucial basis for confirming the authenticity of documents and is also part of the evidence in court. In the insurance field, signature handwriting is used to identify the authenticity of policies and prevent insurance fraud. In the cultural field, signature handwriting reflects the artist's style and personality, holding significant value for in-depth research in

graphology. With technological progress, signature handwriting examination faces many challenges. Signature handwriting is susceptible to factors such as writing habits, emotions, and environment, making the accuracy of handwriting examination complex. Furthermore, the development of signature forgery techniques also brings certain difficulties to the examination work. Considering that the authenticity of most current documents still relies on handwritten signatures for verification, and the cost of manual judgment is too high, there is an urgent need to develop an accurate and efficient signature verification technology.

Signature verification technologies are divided into online signature verification technology and offline signature verification technology based on the input method. For online signature verification, researchers can obtain dynamic information about the signing process, such as stroke trajectories, inclination, and pen pressure [16,31,8,9]. In offline signature verification technology, researchers can only obtain static information, which is signature images captured by scanners or cameras [34,17,1,42]. Because static information provides less information than dynamic information, offline signature verification is more challenging than online signature verification. In today's environment, where paper documents are widely used, offline signature verification has a more widespread application space. Signature verification technology is also divided into writer-dependent and writer-independent methods based on whether it is related to the writer. In writer-dependent methods, researchers' test samples depend on training samples, meaning that each signatory in the test set has a certain amount of signature samples in the training set [21,22,2]. In practical applications, it is impractical to collect and train a large number of samples for each user. In writer-independent methods, the users in the training set and the test set are independent of each other [36,32], thus, they are more valuable in practical applications.

Signature forgery methods are classified into three types based on the proficiency of forgery: random forgery, simple forgery, and skilled forgery [11]. Random forgery signatures have no information about the imitated person, so they differ greatly from genuine samples. Simple forgery involves forged samples that do not follow the writing style of the imitated person, having some similarity to genuine samples. Skilled forgery is performed by professionals who analyze the signature characteristics of the imitated person, resulting in highly similar forged signatures. For skilled forged samples, non-professionals generally cannot distinguish them. Therefore, if criminal organizations obtain relevant information about the imitated person and meticulously forge signatures for criminal activities, this will have adverse effects on the original signatory. Furthermore, for the writer themselves, signatures written in different environments can also vary greatly. Therefore, finding the differences between genuine and forged samples will be a challenging task. To facilitate researchers' study of offline signature verification methods, many public offline signature verification datasets are currently available in academia, such as the English CEDAR dataset [28], GPDS dataset [24], the BHSig260 dataset [7] which includes Bengali and Hindi, and the Chinese MSDS dataset [42], ChiSig dataset [37].

Before the rise of deep learning, researchers typically used traditional image processing methods such as feature matching for signature verification. For example, references [6] and [30] developed the first offline and online signature verification systems; reference [12] utilized the stroke directionality of characters for directional decomposition, then performed band decomposition on the sub-images of each direction, using the decomposed sampled signal values as handwriting features, and employed feature matching methods for writer identification; reference [25] performed identity discrimination through multi-channel two-dimensional Gabor filtering and other methods. Nowadays, researchers are continuously exploring new methods for signature handwriting examination, and with the rise of deep learning and related technologies, reference [3] adopted a Siamese network to extract features from two input sample images separately, and then used metric learning methods to determine the similarity distance between the two signatures, selecting a threshold to determine if they were written by the same person. This metric learning method has significant limitations: on one hand, most metric learning methods use Euclidean distance for calculation, and Euclidean distance only considers the absolute distance between two points, easily overlooking direction and scaling information, and not considering the correlation between data, thus ignoring the relationships between values within feature vectors; on the other hand, its metric threshold is solved through an iterative process, which, although it can obtain the optimal solution for the current dataset, has low generalization ability, and the same threshold will have completely different effects on different datasets. Therefore, reference [4] proposed DeepHSV to address this drawback, using a two-channel discriminative method for offline handwriting verification. By image fusion, two images to be compared are fused into a single image for model input,

which can effectively solve the limitations of metric learning. However, they directly fuse the images before model input, at which point the features of the two compared images are not yet very distinct, leading to the loss of a large number of fine features between different images, thus making it impossible to distinguish meticulously forged signatures. Reference [33] proposed an inverse discriminative octuple attention mechanism, where inverse discriminative images are attached as attention to the original images, making the model focus more on stroke features, and achieved good results on multiple datasets. The limitation of this method is that it focuses too much on the features of the original image and only uses inverse discriminative features as auxiliary judgment information. This paper believes that handwriting features can be obtained not only from the original image but also from inverse grayscale images, which contain a large amount of image features.

Offline handwriting signature verification technology can be regarded as a binary classification task, but it differs significantly from traditional image classification. The differences are: 1) The similarity between the two input images in a handwriting verification system is much higher than in other fields, and the detailed differences between the two images are too sparse; 2) The images are grayscale single-channel images; 3) The essence of handwriting verification system discrimination is style comparison, and improper design can easily lead to overfitting. To address these issues, different scholars have proposed different solutions, such as IDN [33], TransOSV [18], LGR [23], etc. The above methods use CNN or self-attention [33–34] techniques, which are generally classified into different types. However, ACMix proposed in reference [13] proves that the two methods have a strong potential relationship. This paper uses it as the discrimination module of the model, which will make the model focus more on the sparse information features of the fused images to achieve higher discrimination accuracy.

This paper addresses the limitations of two-channel discriminative methods by designing a Multi-channel Feature Fusion Network framework. It employs dual inverse discriminative attention for feature extraction and enhancement of original and inverse grayscale images, integrates the extracted multi-dimensional vectors through channel fusion, and finally uses ACMix for image similarity judgment. This network model has achieved good results on four datasets: CEDAR, BHSig-B, BHSig-H, and ChiSig, demonstrating the effectiveness and generality of the proposed method.

The main contributions of this paper are as follows: 1) Proposed the framework, which enhances the differences between genuine and forged images by fusing multi-Siamese networks to extract multi-dimensional detailed features of input images; 2) Improved the inverse discriminative attention module, strengthening the ability to extract signature features through a dual inverse discriminative attention mechanism; 3) Conducted experiments on CEADR, BHSig-B, BHSig-H, and ChiSig datasets, achieving excellent results superior to baseline papers and most existing methods.

1.1 Siamese Network

Deep metric learning methods primarily involve two samples passing through the same network to generate sample vectors, after which the distance between these two samples is calculated to determine if they belong to the same class. This network is known as a Siamese network. Siamese networks, also called twin networks, are a special neural network structure that can input two images for feature extraction, with the two models sharing weights. In 1993, Siamese networks were first proposed for signature recognition on American checks [18].

Due to their simple structure and ease of implementation, Siamese networks are widely used in image similarity measurement. After passing through the same feature extractor, the extracted features have strong image representativeness. Generally, this network is often used to handle verification problems where the two inputs do not differ significantly. The network takes a pair of samples as input and is trained to make samples with the same label closer in the feature space, and samples with different labels further apart. Therefore, this network has promoted the development of offline signature verification. For example: SigNet proposed in reference [37], MSDN proposed in reference [23], TransOSV proposed in reference [12], etc. The basic network framework of a Siamese network is shown in Figure 1: where A and B are the two input samples, Network 1 and Network 2 are feature extraction networks, and the two networks share parameters. After inputting images, feature vectors a and b are generated through the feature extraction network, and the metric distance between samples a and b is calculated using a metric function. Finally, the network parameters are optimized using a contrastive loss function or other loss functions.

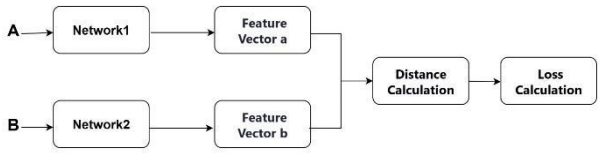

Figure 1: Network Structure Diagram

By generating two feature vectors through a Siamese network, and adhering to the principle that images of the same category are closer and different images are further apart, an optimal threshold can be found by traversing the range between the minimum and maximum distances. However, the method of traversing to find the optimal threshold has a significant limitation: the current threshold is obtained by traversing the current training and test sets, resulting in very low algorithm scalability. Moreover, using Euclidean distance to judge the similarity of different images, while Euclidean distance only considers the absolute distance between two points, easily overlooks information about direction and scaling, and does not consider the correlation between data, thus ignoring the relationships between values within feature vectors. Therefore, a new method is needed to solve this problem.

1.2 Two-Channel Discriminative Network

Another mainstream offline signature verification method is the two-channel discriminative method, which fuses two images and directly outputs 0/1 to determine if they belong to the same class. The biggest difference between this method and the Siamese network is that the Siamese network generates vectors from two samples through the same network structure and then makes a judgment, while the two-channel discriminative network fuses the two images into a single two-channel image before inputting them into the network, and then inputs this single image into a monolithic network to obtain the result of whether they are of the same class. In a two-channel discriminative network, the network does not explicitly extract the input features, but measures their distance in the first step. This design greatly reduces the search parameter space, making two-channel networks particularly suitable for signature verification. The image similarity calculation method based on two channels was proposed by reference [38], and since its proposal, it has achieved considerable results in the field of offline signature verification. For example, reference [6] used two-channel fusion and dual logit output as supervision conditions for training in offline signature verification. Reference [12] proposed an offline signature framework based on two channels and dual Transformers, etc. The basic network diagram of a two-channel discriminative network is shown in Figure 2: where A and B are the two input samples, and the network model is a feature extraction network. After inputting two images, they are first fused into a new image C through image preprocessing before entering the monolithic network. Then, C is input into the monolithic network, and the network output directly indicates whether they were written by the same person, i.e., 0 or 1.

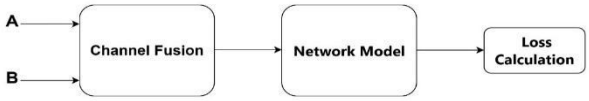

Figure 2: Structure of 2-channel network

Essentially, the two-channel discriminative method treats image similarity judgment as a binary classification method. Through the two-channel discriminative network, the calculation of similarity distance is performed in the first step of the network, and the network directly outputs whether the signatures were written by the same person. Compared to the Siamese network method, this method significantly reduces the search parameter space, effectively speeding up network training; on the other hand, the method of directly outputting results avoids the limitations of the Siamese network's threshold, and the accuracy will not be significantly affected when changing training datasets or adding data. Current networks for two-channel discriminative methods directly perform fusion on original images or after image cropping, i.e., measuring the distance on the initial two images. At this point, the image features are not yet obvious, and simply fusing them will result in the loss of a large number of fine features, ultimately leading to poor model performance.

1.3 Grayscale Processing

In offline signature verification, this paper inputs two single-channel images. Reference [14] also attempted to train with three-channel color images, but the effect was not as good as grayscale images. In grayscale images, different grayscale distributions will have a significant impact on the model's results. For example, black text on a white background and white text on a black background, different inputs will have a significant impact on the training of the same model. This is because in signature verification images, the data model only needs the feature information of the handwriting strokes, and most background information is invalid or even harmful. If the background information consists of pixels with a value of 0, the result after convolution will not change, which has a considerable impact on feature extraction and even the model's output. However, this does not mean that white-on-black images are all invalid information; they also contain detailed and important information. Addressing this issue, reference [30] proposed an inverse discriminative network, where the network input is a black-on-white image. This network enhances the effective information for signature

verification through grayscale processing and a multi-path attention module. The attention module of this method extracts features from inverse grayscale images and creates an attention module loaded onto the original grayscale images, making the model focus more on the stroke information of the image. This method innovatively extracts features from both black-on-white and their inverse grayscale images. However, the focus of feature extraction in this method is on the original grayscale image, neglecting that its inverse grayscale image is not only a tool for auxiliary attention but also contains a large amount of handwriting stroke information.

1.4 ACMix

Convolutional kernels and self-attention are two powerful techniques for representation learning, and there is a strong potential relationship between them because most of the computations in these two paradigms are actually performed through the same operations. Specifically, a convolution with kernel size k × k can be divided into k2 individual 1 × 1 convolutions, followed by shifting and summing operations. In ACMix, 1 × 1 convolutional kernels are first used to project input features into queries, keys, and values, and then the attention weights and the aggregation of value matrices, i.e., the aggregation of local features, are calculated. Therefore, ACMix can elegantly integrate these two seemingly different paradigms, enjoying the benefits of both self-attention and convolution, while having smaller overhead compared to pure convolution or self-attention [33].

This paper proposes a network structure to address the feature loss problem in two-channel discriminative networks. It employs a quadruple Siamese network and a dual inverse discriminative attention mechanism for feature extraction, fuses the extracted multiple subtle features through channel fusion, and finally uses a combination of self-attention and convolutional networks to determine whether it is a genuine sample pair.

2. METHODOLOGY

As an end-to-end signature verification system, it consists of feature extraction, channel fusion, and an ACMix module. A pair of signature images first undergo inverse grayscale acquisition, generating a total of four images, which are then input into a quadruple Siamese network. Then, the grayscale and inverse grayscale images of the same image are weighted and calculated through a dual inverse discriminative attention module to extract a large number of detailed features. Finally, the extracted different image representations are fused through channel fusion, and a combination of convolutional neural networks and self-attention is used for discriminative processing to achieve high-similarity image discrimination.

2.1 Dual Inverse Discriminative Attention Module

The feature extractor of this network adopts a quadruple Siamese network structure. This network consists of two convolutional blocks, each containing two convolutional layers activated by ReLU function. Each convolutional layer has a size of 3 × 3, stride of 1, and padding of 1. The dimension of each convolutional block is 64, 128. The reference image and its inverse grayscale image, and the test image and its inverse grayscale image are respectively input into the feature extraction network, and the networks share weights. Between the grayscale image convolutional block and the corresponding inverse grayscale image convolutional block, four dual stroke attention modules are connected. Each attention module connects the convolutional module in the discriminative flow and the convolutional module in the inverse flow, as shown in Figure 3.

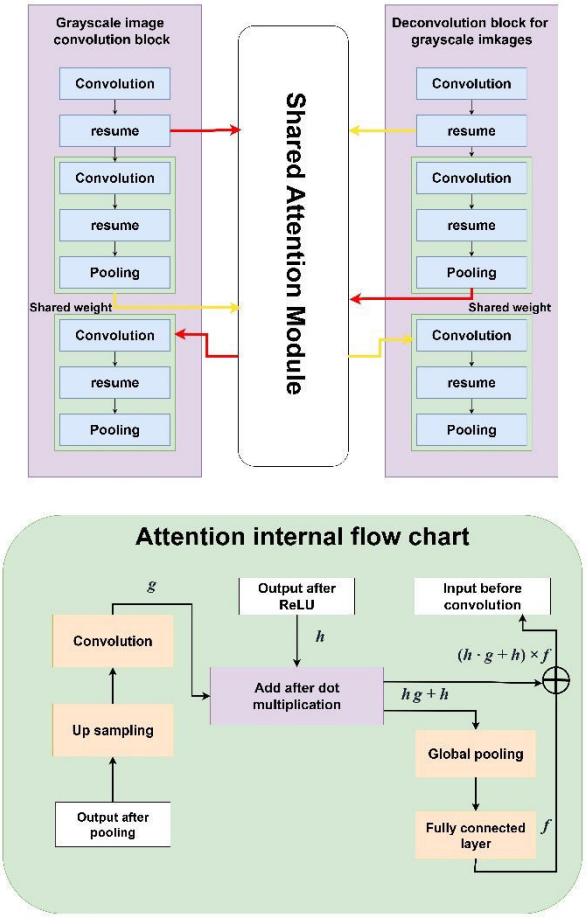

Figure 3: Dual reverse forensic attention module

The grayscale image convolutional block and the inverse grayscale image convolutional block of the same image are simultaneously input into the dual inverse discriminative attention module. Among them, according to the different moments of output from the grayscale image convolutional block and the inverse grayscale image convolutional block, two data streams enter the shared attention. The internal flowchart of the attention module is shown in the right part of Figure 3, taking the yellow data stream in the left part as an example: The feature vector output from the grayscale image convolutional block is input into the upsampling structure, which uses the nearest neighbor algorithm for upsampling and performs convolution operations with Sigmoid activation, outputting g. The inverse grayscale image convolutional block outputs h after ReLU. h is multiplied by the elements of g, and then h is added to produce the intermediate attention measurement h · g + h, where "·" denotes element-wise multiplication. The subsequent Global Average Pooling (GAP) layer and a fully connected layer (FC) with Sigmoid activation receive the intermediate attention measurement and output a weight vector f. Each channel is multiplied by each element of f to generate the final attention (h · g + h) × f, which is then output to the second layer of the inverse grayscale image convolutional block for convolutional processing. This method has two data streams, red and yellow, depending on the input and output, and the shared attention module parameters are shared between them. The dashed boxes in the two convolutional blocks also share weights, and the two convolutional blocks also share weights. Assuming the ReLU output of the grayscale image is x1, and the output of the grayscale image convolutional block is y1; the ReLU output of the inverse grayscale image is x2, and the output of the inverse grayscale image convolutional block is y2; the shared convolutional blocks in the dashed part are collectively referred to as w. Therefore, different data streams have different formulas, specifically as shown in equations (1) to (4):

out1 = h(x1) g[w(x2)] + h(x1) (1) y1 = w[out1 × f (out1)] (2) out2 = h(x2) · g[w(x1)] + h(x2) (3) y2 = w[out2 × f (out2)] (4)

Among them, equations (1) and (2) are for the red data stream, and equations (3) and (4) are for the yellow data stream.

In the attention module, this paper processes grayscale images and inverse grayscale images separately for attention, forming dual inverse discriminative attention.. By comparing with IDN's attention, it is found that although IDN has quadruple attention, as the convolution operation deepens, the focus of the attention module becomes more abstract. In contrast, the framework's attention module, on one hand, introduces dual features, creating constraints between attentions, enabling accurate focus on stroke edge information even in the second layer of attention feature maps; on the other hand, by reducing the number of layers, it extracts sufficiently detailed features during channel fusion. Through the multi-path attention mechanism, the important features for signature verification are enhanced.

Since the attention module in this paper connects the original grayscale image and the inverse grayscale image, the final attention mask will guide the network to learn discriminative features for signature verification and suppress misleading information. The entire framework has 4 attention modules connecting different convolutional modules, applying the attention mechanism at different scales and resolutions.

2.2 Multi-channel Fusion

The two-channel discriminative method fuses two input single-channel images into a two-channel image and directly outputs whether they are similar to quickly obtain results. However, because the two-channel discriminative network performs channel fusion in the original state of the images, the image features are not yet obvious, and simply fusing them will result in the loss of a large number of fine features, ultimately leading to poor model performance. Therefore, in the framework, four images, each with 128-dimensional feature information after feature extraction, are fused, totaling 512-dimensional feature information. Compared to traditional two-channel discriminative networks, our framework not only includes the convolutional features of the two images to be discriminated but also includes the convolutional features of the inverse grayscale images of the two images. This method considers more channel information during the fusion process, allowing the network to capture more information and increasing the diversity of fused features. Therefore, in the subsequent discrimination stage, this network can achieve higher accuracy.

The proposed framework connects the extracted features of the reference image, reference image inverse grayscale image, test image, and test image inverse grayscale image. Each image has 128 dimensions. After passing through the discriminative module, it finally outputs 0/1 to determine whether they are the same person. Compared to the two-channel discriminative method, the multi-channel image formed by multi-

channel fusion has more fused image features. The distance calculation between images changes from one positive and one negative sample, totaling two images, to 256 dimensions for each positive and negative sample to calculate the difference. Because there are more image features, the calculated distance is more accurate.

2.3 Discriminative Module

In the discriminative module, this paper primarily uses the ACMix module, supplemented by two small convolutional blocks for discrimination. After feature fusion, a 512-dimensional feature vector is obtained, with many and large differences between features. To accurately extract features, the ACMix module, which combines convolution and self-attention, is used for feature extraction.

The 512-dimensional feature representation after channel fusion is not directly input into a fully connected layer for classification. Instead, it first passes through a discriminative module based on a monolithic network model [33]. This module performs overall feature learning and judgment based on self-attention and convolutional networks, finally outputting a 0/1 binary classification result. The structure of the discriminative module consists of two small convolutional modules and one ACMix module. The input of the first small convolutional module is the integrated 512-dimensional features after channel fusion. Then, the ACMix's convolution and self-attention mechanisms are used for feature extraction. Finally, a small convolutional module is used for feature summarization, and then it enters a multi-layer perceptron for classification. At this point, the features entering the multi-layer perceptron are the 512-dimensional image features extracted by the discriminative network, which contain the overall difference information between the reference image, reference image inverse grayscale image, test image, and test image inverse grayscale image.

Global average pooling is introduced in the multi-layer perceptron to reduce network redundancy. To avoid overfitting, this paper uses 0.5 Dropout. Finally, the entire network will output a Sigmoid-activated feature value, generating a judgment probability between 0 and 1. In the accuracy judgment process, this paper sets a probability less than or equal to 0.5 as a forged signature, and a probability greater than 0.5 as a genuine signature. The loss function uses binary cross-entropy loss, and its formula is:

$$L = -(1/n) \sum[yi \lg(pi) + (1 - yi)\lg(1 - pi)] \quad (5)$$

Where yi represents the true label of sample i, 1 for positive class, 0 for negative class. pi represents the probability that sample i is predicted as positive, and similarly, 1 − pi is the probability that the sample is predicted as negative.

3. EXPERIMENT

The quantity and quality of datasets have a significant impact on the model. Currently, with the in-depth research of domestic and foreign scholars in the field of offline handwriting verification, many public offline datasets have been proposed. This paper will use the English CEDAR dataset, the BHSig260 dataset (including Bengali and Hindi), and the Chinese ChiSig dataset for model testing and evaluation. Statistical information for various datasets is shown in Table 1.

The CEDAR dataset is a signature sample dataset in English. It consists of samples from 55 signers, with each signer having 24 genuine signature samples and 24 forged signature samples. According to previous work, this paper selects samples from 50 individuals for training and samples from the remaining 5 signers for testing. For each signer, this dataset has 276 reference-genuine sample pairs and 576 reference-forged sample pairs.

Table 1: Offline signature verification dataset

Data set name	Language	Signature type	Number of pictures	Real to fake sample ratio
CEDAR	English	55	2624	24/24
BHSig-B	Bengali	100	5400	24/30
BHSig-H	Hindi	160	8640	24/30
ChiSig	Chinese	102	10242	-/-

To ensure a balance of positive and negative samples, this paper will randomly draw reference-forged sample pairs based on the number of reference-genuine sample pairs. Therefore, for each signer, this paper will have 276 reference-genuine sample pairs and 276 reference-forged sample pairs for training and testing.

The BHSig260 dataset includes Bengali and Hindi datasets, which are treated as two different datasets in this paper. The BHSig-B dataset contains Bengali signature images from 100 signers. Each signer has 24

genuine signatures and 30 forged signatures. Based on previous experience, this paper randomly selects signatures from 50 signers for training, and signatures from the remaining signers for testing. The BHSig-H dataset contains Hindi signature images from 160 signers. Each signer has 24 genuine signatures and 30 forged signatures. Similarly, this paper will randomly select signatures from 100 signers as the training set to train the model, and signatures from the remaining 60 signers as test data. For each signer in the above two datasets, this paper also randomly draws 276 reference-genuine sample pairs and 276 reference-forged sample pairs for training and testing.

Reference [11] constructed a novel Chinese document offline signature forgery detection benchmark dataset, ChiSig, which includes all tasks such as signature detection, restoration, and verification. The dataset consists of clean handwritten signatures, synthetically interfered handwritten signatures, and synthetic documents with handwritten signatures. The authors randomly generated 500 names and then asked volunteers to sign according to certain rules to obtain clean signature data, which can be used for signature verification tasks. Because the number of volunteers is greater than the number of names, there are cases where different writers have the same name, which poses a great challenge for signature verification. Afterwards, the authors obtained scanned documents that can be used as synthetic backgrounds from public resources such as the XFUND dataset, Chinese national standards, and patents. For this dataset, this paper randomly draws 250 signatures as the training set and 250 signatures as the test set. For each name, signatures written by the same volunteer are treated as genuine sample pairs, and signatures written by different volunteers are treated as forged sample pairs. For dedicated forged data, they are only treated as forged sample pairs, and forged data are not treated as genuine sample pairs. To ensure data balance between genuine and forged sample pairs, this paper removes redundant sample pairs.

3.2 Evaluation Metrics

For the CEDAR and BHSig260 datasets, this paper will follow the settings in reference [30] and use False Rejection Rate (FRR), False Acceptance Rate (FAR), and Accuracy (ACC) to comprehensively evaluate the framework and compare it with other existing methods.

FRR is defined as the ratio of the number of false rejections to the number of genuine samples. FAR is defined as the ratio of the number of false acceptances to the number of forged samples. ACC is defined as the ratio of the number of correctly judged samples to the total number of samples.

For the ChiSig dataset, this paper uses the evaluation metrics proposed by the dataset authors: Accuracy, Equal Error Rate (EER), and True Acceptance Rate (TAR) for comparison. EER evaluates the balance point where FRR equals FAR; the lower the EER, the better the model performance. The calculation method for TAR is shown in equations (6) to (8), and TAR is only calculated when the False Acceptance Rate (FAR) equals 10−3:

FAR = (Number of False Acceptances) / (Number of Forgeries) (6) FRR = (Number of False Rejections) / (Number of Genuine Samples) (7) TAR = 1 − FRR (8)

3.3 Comparative Experiments

To verify the model's effectiveness, this paper selects the latest deep learning models for comparison based on the current development of handwriting verification tasks, namely SigNet (2017arXiv) [37], IDN (2019CVPR) [30], DeepHSV (2019ICDAR) [6], SDINet (2021AAAI) [13], SURDS (2022ICPR) [39], 2C2S (2023EAAI) [40], TransOSV (2022ICME) [12]. These models include methods combining Siamese networks with metric learning, as well as methods using two-channel discrimination. The comparison results are sufficient to illustrate the advantages of the proposed model proposed in this paper. For convenience of observation, the optimal solution is bolded, the suboptimal solution is underlined, and the third best solution is wavy. CEDAR, BHSig-B, and BHSig-H are shown in Table 2, Table 3, and Table 4, respectively. The results for the ChiSig dataset will be introduced in Section 3.4.

In the experimental results on the CEDAR dataset, the proposed model achieved 100% accuracy. The main reason is that this dataset has a small number of samples, a simple structure, and large differences, so many methods have achieved good results on this dataset. Comprehensive analysis shows that model's ACC improved by 3.62% and 1.75% compared to IDN and SDI, respectively, and achieved 100% like SigNet, DeepHSV, and 2C2S. In the BHSig-B dataset, the experimental results show that the model has a greater advantage than current mainstream offline handwriting verification algorithms, achieving an accuracy of 95.61%, and this is also proven in the comparison of FRR and FAR, reaching optimal or suboptimal. Compared to IDN, model's ACC improved by 0.29%. Compared to the latest algorithms 2C2S and TransOSV, it improved by 2.36% and 5.56%, respectively. This is

sufficient to prove the superiority of the model proposed in this paper.

Table 2: Comparison on CEDAR dataset (%)

Model name	FRR	FAR	ACC
SigNet (2017arXiv)	0	0	100.00
DeepHSV (2019ICDAR)	-	-	100
IDN (2019CVPR)	2.17	5.87	96.38
SDINet (2021AAAI)	3.42	0.73	98.25
2C2S (2023EAAI)	0	0	100.00
OURS	0	0	100.00

Table 3 Comparison on BHSig-B dataset (%)

Model Name	FRR	FAR	ACC
SigNet (2017arXiv)	13.89	13.89	86.11
DeepHSV (2019ICDAR)	—	—	88.08
IDN (2019CVPR)	5.24	4.12	95.32
SDINet (2021AAAI)	7.86	3.30	94.42
SURDS (2022ICPR)	5.42	19.89	87.34
2C2S (2023EAAI)	8.11	5.37	93.25
TransOSV (2022ICME)	9.95	9.95	90.05
OURS	3.86	3.84	95.61

Table 4 Comparison on BHSig-H dataset (%)

Model Name	FRR	FAR	ACC
SigNet (2017arXiv)	15.36	15.36	84.64
DeepHSV (2019ICDAR)	—	—	86.66
IDN (2019CVPR)	4.93	8.99	93.04
SDINet (2021AAAI)	3.77	6.24	95.00
SURDS (2022ICPR)	8.98	12.01	89.50
2C2S (2023EAAI)	9.98	8.66	90.68
TransOSV (2022ICME)	3.39	3.39	96.61
OURS	4.89	4.89	95.70

Similar to CEDAR and BHSig-B, our model also achieved good results on the BHSig-H dataset. Compared to the latest algorithms, model achieved an accuracy of 95.7% on the BHSig-H dataset, although it is not the optimal result, its FRR is third best, and the others are suboptimal. Furthermore, compared to the optimal, model's accuracy is only 0.89% lower, while in BHSig-B, compared to the optimal model TransOSV in BHSig-H, our model achieved a 5.56% lead in accuracy. This is sufficient to show that model's generalization ability is superior to TransOSV.

3.4 Ablation Experiment

In addition, this paper conducted ablation experiments on the ChiSig dataset. InceptionResnet is the baseline model provided in the dataset paper [11]. This paper conducted comparative experiments by reproducing SigNet and IDN code.

As shown in Table 5, the baseline IDN compared with its channel fusion method, the channel fusion method improved the accuracy by 0.9% compared to the original method; the dual inverse discriminative attention expanded the information of the inverse grayscale image, providing more detailed information during channel fusion, which improved the accuracy to 88.96%, an increase of 3.24% compared to channel fusion. The ACMix discriminative structure further improved the model's accuracy to 95.23%.

Table 5 Ablation experiment on ChiSig dataset (%)

Model Name	EER	TAR	ACC
InceptionResnet	6.60	28.10	93.60
SigNet	—	—	82.28
IDN (Baseline)	17.91	10.50	84.82
IDN (Channel Fusion)	14.81	9.61	85.72
IDN (Channel Fusion + Attention)	11.38	7.82	88.96
OURS (No Inverse Gray, No Attention)	11.78	32.49	88.09
OURS (No Inverse Gray, Single Attention)	10.83	—	89.20
OURS (Inverse Gray, No Attention)	7.84	—	92.14
OURS (Full Model)	5.19	28.96	95.23

To demonstrate the impact of inverse grayscale images and corresponding attention on the results, this paper

also conducted experiments by removing grayscale images and attention. 'No inverse grayscale image' means the model only inputs reference images and test images. 'Single attention' means that in the dual attention module, the input for dot product and upsampling is provided by itself, and everything else is consistent with the final model.

Table 6 Main parameters on ChiSig dataset (%)

Model Name	FRR	FAR	ACC	Notes
IDN (Baseline)	10.46	17.91	84.82	Original implementation
IDN (Channel Fusion)	9.61	18.97	85.72	+Feature combination
IDN (Channel Fusion + Attention)	7.82	14.27	88.96	+Attention mechanism
OURS (No Grayscale Inversion, No Attention)	21.91	17.26	88.09	Basic version
OURS (No Grayscale Inversion, Single Attention)	15.59	16.30	89.20	+Attention layer
OURS (Grayscale Inversion, No Attention)	6.90	17.18	92.14	+Image preprocessing
OURS (Full Model)	5.34	5.34	95.23	Complete configuration

For no inverse grayscale image, after introducing single attention, the accuracy increased by 1.11%, while introducing inverse grayscale images increased the accuracy by 4.05%. Experimental results show that the addition of attention and inverse grayscale images is feasible, and the addition of inverse grayscale images has a greater improvement effect than the addition of attention.

This ablation experiment proves the rationality of the proposed method. In addition, to facilitate future researchers to compare using FRR and FAR metrics, this paper also calculated the FRR and FAR metrics of our proposed model on the ChiSig dataset, as shown in Table 6.

3.5 Cross-Language Experiment

Furthermore, this paper also conducted cross-language tests. In this work, CEDAR, BHSig-B, BHSig-H, and ChiSig, four different languages, were used for testing. This paper trained the model using the training set of one language and tested it on the training sets of the remaining languages. For example, this paper trained the model on the BHSig-B training dataset and tested it on the BHSig-H test dataset. The division of training and test data is the same as in the experiments on each independent dataset. Table 7 shows the accuracy of cross-language tests, where rows correspond to training languages and columns correspond to test languages.

Table 7 shows that cross-language signature verification performance rapidly declines. This paper believes that the essence of an offline signature verification system is style feature matching.

Each person's signature is closely related to their writing style habits, and different language styles have different writing habits, leading to the inability of the current dataset's learned style to be applied to other datasets. The accuracy of the BHSig-B dataset and BHSig-H dataset is higher than other datasets, possibly because the writing styles of Hindi and Bengali are more similar.

Table 7 Cross-language test (%)

Training Set → Test Set	CEDAR	BHSig-B	BHSig-H	ChiSig
CEDAR	100.00	48.76	49.89	57.48
BHSig-B	64.86	95.61	82.79	63.71
BHSig-H	50.11	86.27	95.70	20.00
ChiSig	54.60	70.02	55.37	95.23

4. CONCLUSION

This paper proposes a novel offline handwriting verification model, for handwritten signature verification in writer-independent scenarios. This model first extracts features through two layers of convolutional networks and a dual attention module, then performs feature fusion through channel fusion, and finally uses the ACMix discriminative module to determine the similarity of multiple images. It uses an inverse supervision mechanism and a dual attention mechanism to solve the problem of insufficient detailed feature information in traditional channel fusion methods. In testing, by inputting reference signature images and test signature images, the model directly outputs whether the test signature is genuine or forged. Experimental results demonstrate the advantages and potential of the proposed method. Future work will focus on research into cross-language signature verification and recognition.

REFERENCES

1. Soleimani A, Araabi B N, Fouladi K. Deep multitask metric learning for offline signature

verification. Pattern Recognition Letters, 2016, 80: 84−90

2. Ferrer M A, Alonso J B, Travieso C M. Offline geometric parameters for automatic signature verification using fixed-point arithmetic. IEEE Transactions on Pattern Analysis And Machine Intelligence, 2005, 27(6): 993−997
3. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, et al. Attention is all you need. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc., 2017. 6000−6010
4. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X H, Unterthiner T, et al. An image is worth 16×16 words: Transformers for image recognition at scale. In: Proceedings of the 9th International Conference on Learning Representations. Austria: OpenReview.net, 2021.
5. A Transformer Based Handwriting Recognition System Jointly Using. arXiv, 2025.
6. Li C, Lin F, Wang Z Y, Yu G, Yuan L, Wang H Q. DeepHSV: User-independent offline signature verification using two-channel CNN. In: Proceedings of the International Conference on Document Analysis and Recognition (ICDAR). Sydney, Australia: IEEE, 2019. 166−171
7. Liu Cheng-Lin, Liu Ying-Jian, Dai Ru-Wei. Writer identification by multichannel decomposition and matching. Acta Automatica Sinica, 1997, 23(1): 56−63
8. Hafemann L G, Oliveira L S, Sabourin R. Fixed-sized representation learning from offline handwritten signatures of different sizes. International Journal on Document Analysis and Recognition (IJDAR), 2018, 21(3): 219−232
9. Hafemann L G, Sabourin R, Oliveira L S. Learning features for offline handwritten signature verification using deep convolutional neural networks. Pattern Recognition, 2017, 70: 163−176
10. Xia X H, Song X Y, Luan F G, Zheng J G, Chen Z L, Ma X F. Discriminative feature selection for on-line signature verification. Pattern Recognition, 2018, 74: 422−433
11. Yan K H, Zhang Y, Tang H R, Ren C K, Zhang J, Wang G A, et al. Signature detection, restoration, and verification: A novel Chinese document signature forgery detection benchmark. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). New Orleans, USA: IEEE, 2022. 5163−5172
12. Li H, Wei P, Ma Z Y, Li C K, Zheng N N. Offline signature verification with transformers. In: Proceedings of the IEEE International Conference on Multimedia and Expo (ICME). Taipei, China: IEEE, 2022. 1−6
13. Li H, Wei P, Hu P. Static-dynamic interaction networks for offline signature verification. In: Proceedings of the 35th AAAI Conference on Artificial Intelligence. Vancouver, Canada: AAAI Press, 2021. 1893−1901
14. Bhattacharya I, Ghosh P, Biswas S. Offline signature verification using pixel matching technique. Procedia Technology, 2013, 10: 970−977
15. Bromley J, Bentz J W, Bottou L L, Guyon I, Lecun Y, Moore C, et al. Signature verification using a “Siamese” time delay neural network. International Journal of Pattern Recognition and Artificial Intelligence, 1993, 7(4): 669−688
16. Hu J, Chen Y B. Offline signature verification using real adaboost classifier combination of Pseudo-dynamic features. In: Proceedings of the 12th International Conference on Document Analysis and Recognition. Washington, USA: IEEE, 2013. 1345−1349
17. Xing Z J, Yin F, Wu Y C, Liu C L. Offline signature verification using convolution Siamese network. In: Proceedings of SPIE 10615, 9th International Conference on Graphic and Image Processing (ICGIP). Qingdao, China: SPIE, 2017. 415−423
18. Bromley J, Guyon I, LeCun Y, Säckinger E, Shah R. Signature verification using a “Siamese” time delay neural network. In: Proceedings of the 6th International Conference on Neural Information Processing Systems. Denver, Colorado: Morgan Kaufmann Publishers Inc., 1993. 737−744
19. Zou Jie, Sun Bao-Lin, Yu Jun. Online handwriting matching algorithm based on stroke features. Acta Automatica Sinica, 2016, 42(11): 1744−1757
20. Cpałka K, Zalasiński M, Rutkowski L. New method for the online signature verification based on horizontal partitioning. Pattern Recognition, 2014, 47(8): 2652−2661
21. Jain A K, Ross A, Prabhakar S. An introduction to biometric recognition. IEEE Transactions on Circuits and Systems for Video Technology, 2004, 14(1): 4−20
22. Kalera M K, Srihari S, Xu A H. Offline signature verification and identification using distance statistics. International Journal of Pattern Recognition and Artificial Intelligence, 2004, 18(7): 1339−1360
23. Liu L, Huang L L, Yin F, Chen Y B. Offline signature verification using a region based deep metric learning network. Pattern Recognition, 2021, 118: Article No. 108009
24. Herbst N M, Liu C N. Automatic signature verification based on accelerometry. IBM Journal of Research and Development, 1977, 21(3): 245−253
25. Cairang X M, Zhaxi D J, Yang X L, Hou Y, Zhao Q J, Gao D G, et al. Learning generalisable representations for offline signature verification. In: Proceedings of the International Joint

Conference on Neural Networks (IJCNN). Padua, Italy: IEEE, 2022. 1−7

26. Enhancing Signature Verification Using Triplet Siamese Similarity. MDPI, 2024.
27. Combining Multi-Scale Fusion and Attentional Mechanisms for. MDPI, 2025.
28. Nagel R N, Rosenfeld A. Steps toward handwritten signature verification. In: Proceedings of the 1st International Joint Conference on Pattern Recognition. 1973. 59−66
29. Learning features for offline handwritten signature verification using. Nature, 2025.
30. Wei P, Li H, Hu P. Inverse discriminative networks for handwritten signature verification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE, 2019. 5764−5772
31. Kumar R, Sharma J D, Chanda B. Writer-independent off-line signature verification using surroundedness feature. Pattern Recognition Letters, 2012, 33(3): 301−308
32. Zhang P R, Jiang J J, Liu Y L, Jin L W. MSDS: A large-scale Chinese signature and token digit string dataset for handwriting verification. In: Proceedings of the 36th International Conference on Neural Information Processings Systems. New Orleans, USA: 2022. 36507−36519
33. Pan X R, Ge C J, Lu R, Song S J, Chen G F, Huang Z Y, et al. On the integration of self-attention and convolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, USA: IEEE, 2022. 815−825
34. Okawa M. Synergy of foreground-background images for feature extraction: Offline signature verification using Fisher vector with fused KAZE features. Pattern Recognition, 2018, 79: 480−489
35. A Survey of Offline Handwriting Signature Verification. ResearchGate, 2025.
36. Pal S, Alaei A, Pal U, Blumenstein M. Performance of an offline signature verification method based on texture features on a large Indic-script signature dataset. In: Proceedings of the 12th IAPR workshop on Document Analysis Systems (DAS). Santorini, Greece: IEEE, 2016. 72−77
37. Dey S, Dutta A, Toledo J I, Ghosh S K, Llados J, Pal U. SigNet: Convolutional Siamese network for writer independent offline signature verification. arXiv preprint arXiv: 1707.02131, 2017.
38. Zagoruyko S, Komodakis N. Learning to compare image patches via convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, USA: IEEE, 2015. 4353−4361
39. Chattopadhyay S, Manna S, Bhattacharya S, Pal U. SURDS: Self-supervised attention-guided reconstruction and dual triplet loss for writer independent offline signature verification. In: Proceedings of the 26th International Conference on Pattern Recognition (ICPR). Montreal, Canada: IEEE, 2022. 1600−1606
40. Ren J X, Xiong Y J, Zhan H J, Huang B. 2C2S: A two-channel and two-stream transformer based framework for offline signature verification. Engineering Applications of Artificial Intelligence, 2023, 118: Article No. 105639
41. Guerbai Y, Chibani Y, Hadjadji B. The effective use of the one-class SVM classifier for handwritten signature verification based on writer-independent parameters. Pattern Recognition, 2015, 48(1): 103−113
42. Zhu Yong, Tan Tie-Niu, Wang Yun-Hong. Writer identification based on texture analysis. Acta Automatica Sinica, 2001, 27(2): 229−234
43. EID, A. A., Miled, A. B., Mahmoud, A. F., Abdalla, F. A., Jabnoun, C., Dhibi, A., ... & Belhaj, S. (2024). Leveraging Arabic Text Embedded in Images: Challenges and Opportunities in NLP Analysis. Journal of Intelligent Systems and Applied Data Science, 2(1).