applied sciences

MDPI

Article

DDS: Deepfake Detection System through Collective Intelligence and Deep-Learning Model in Blockchain Environment

Nakhoon Choi and Heeyoul Kim *

Department of Computer Science, Kyonggi University, Suwon-si 16227, Republic of Korea
* Correspondence: heeyoul.kim@kgu.ac.kr

Abstract: With the spread of mobile devices and the improvement of the mobile service environment, the use of various Internet content providers (ICPs), including content services such as YouTube and video hosting services, has increased significantly. Video content shared in ICP is used for information delivery and issue checking based on accessibility. However, if the content registered and shared in ICP is manipulated through deepfakes and maliciously distributed to cause political attacks or social problems, it can cause a very large negative effect. This study aims to propose a deepfake detection system that detects manipulated video content distributed in video hosting services while ensuring the transparency and objectivity of the detection subject. The detection method of the proposed system is configured through a blockchain and is not dependent on a single ICP, establishing a cooperative system among multiple ICPs and achieving consensus for the common purpose of deepfake detection. In the proposed system, the deep-learning model for detecting deepfakes is independently driven by each ICP, and the results are ensembled through integrated voting. Furthermore, this study proposes a method to supplement the objectivity of integrated voting and the neutrality of the deep-learning model by ensembling collective intelligence-based voting through the participation of ICP users in the integrated voting process and ensuring high accuracy at the same time. Through the proposed system, the accuracy of the deep-learning model is supplemented by utilizing collective intelligence in the blockchain environment, and the creation of a consortium contract environment for common goals between companies with conflicting interests is illuminated.

Keywords: blockchain; deepfake detection; collective intelligence; Hyperledger Fabric

Citation: Choi, N.; Kim, H. DDS: Deepfake Detection System through Collective Intelligence and Deep-Learning Model in Blockchain Environment. *Appl. Sci.* **2023**, *13*, 2122. https://doi.org/10.3390/app13042122

Academic Editors: Hui Li, Shancang Li and Konstantinos Demertzis

Received: 8 January 2023
Revised: 1 February 2023
Accepted: 3 February 2023
Published: 7 February 2023

1. Introduction

Currently, Internet content providers (ICPs), such as video hosting services like YouTube, have developed rapidly through the Fourth Industrial Revolution, the spread of smartphones, high levels of informatization, and the improvement of the mobile service environment. Due to the development of ICP, video media and content are widely shared, and as video content is consumed by many users, social influence through video is growing [1]. Deepfake videos with manipulated content or characters, on the other hand, can be used to cause social confusion or political attacks. If they spread through indiscriminate sharing, the resulting damage is uncontrollable [2]. ICPs such as YouTube and Facebook, which are currently widely used, are censoring video content in their own way. However, to guarantee freedom of expression, ICP only responds to content copyright infringement and does not respond to manipulated images. In the era of hyperconnectivity society due to the Fourth Industrial Revolution, social confusion due to manipulated video content, such as deepfake videos, will accelerate. To prevent this, deepfake detection technology through a fair and reliable detection entity is required.

Most of the existing studies to detect deepfake images have focused on determining whether a single image has been manipulated in a local environment through an independent deep-learning model. As a study to detect deepfake images in the form of a platform

or system, the study by Li et al. [3] developed an independent platform to provide the result by determining whether manipulation was performed through the deep-learning model of the service server for the videos that users individually accessed and registered. However, this has limitations in that the user must fully trust the detection subject and must own and upload the original video. Hasan [4] and Patil [5] suggested a plan to integrate and manage video content registered in various services such as YouTube. The above study used a blockchain as a database to find manipulated or unauthorized content through video metadata-based use paths and revision history tracking. However, this assumes an environment in which the original video of the manipulated content is pre-registered and has limitations in tracking the modified items in the original.

This study aims to propose a system in which various ICPs and users mutually cooperate to detect deepfakes and ensure fair and reliable detection results and detection subjects. The proposed system integrates the collective intelligence-based voting and prediction results of deep-learning models in a blockchain environment for fair deepfake detection results. In addition, the detection system of each ICP that performs deepfake detection through a blockchain is flexibly integrated, preventing the decision-making authority of the detection result from being monopolized. The detection process, results, and detection model information are recorded on the blockchain to ensure integrity and persistence. In the proposed system, the fairness of the detection results is supplemented by adding a voting process based on collective intelligence to the deepfake decision process. Collective intelligence-based user voting, which is integrated through voting weights set based on users' reputations, is integrated with the results of the deep-learning model to supplement the accuracy of the final detection results. The main contributions of this study are summarized as follows:

1. In most studies, fairness and transparency of the detecting subject have not been considered important in building a system for detecting deepfakes, degrading the reliability of the system and detection results. This study uses blockchain technology to improve the fairness and stability of deepfake detection subjects by conducting a detection process based on an automated and verifiable smart contract. Along with the detection results, the main information generated during the detection process is recorded in the blockchain, securing the transparency and reliability of the detection results through the integrity of the blockchain. The proposed system forms cooperation between hostile ICPs through the blockchain to derive a common consensus, and ICPs will establish a consortium for pan-service deepfake detection with the blockchain network as a notary.
2. This study aims to propose an integrated voting method that mixes the deep-learning-based detection results of the ICP consortium with the detection results obtained through collective intelligence voting to improve the accuracy of deepfake detection results. Each ICP independently builds a deep-learning-based detection model to derive detection results, and the results of several ICPs are aggregated to produce the detection results of the consortium. Then, random users are selected and voted on to derive deepfake detection results based on collective intelligence. These two detection results are combined to produce the final detection result. This proposed method has higher accuracy and stability than individual deepfake detection models.
3. To improve the accuracy of the collective intelligence voting process based on user participation in a blockchain environment, this study introduced a user reputation management and compensation system, referring to several previous studies. Each user's reputation is continuously updated by reflecting the user's past behavior, and weight based on this reputation is applied to the user's voting results. In addition, the reward system for friendly behavior in the system induces the normal behavior of the voting group and supplements the accuracy of the final voting result.
4. In this study, several recently announced deepfake detection models were applied to independent nodes to implement prototypes of the proposed system and conduct experiments. The experimental results show that the proposed system outperforms

existing single deepfake detection models in terms of accuracy. In addition, the integrated voting method used in the proposed system provides high stability.

In this paper, we propose a deepfake detection system using collective intelligence and deep learning in a blockchain environment. The proposed system integrates the prediction results of independent deep-learning nodes to solve the problem of detection reliability of individual independent detectors, which is a problem of previous research [3]. Through the ensemble of independently operated deep-learning model nodes, each node pursues complementary model development and has a node weight that complements the accuracy and reliability of detection results. Another study [4,5] used a blockchain to check and track content forgery and tampering, but its limitations are clear because it is only detection based on metadata. Through the blockchain, we establish an organic cooperation system of service providers-blockchain network-users. We do not simply use the blockchain as a database, but we derive cooperation for meaningful agreement and detection in mutually hostile business relationships. In this paper, we design a detection cooperative system based on deep learning and collective intelligence in a blockchain environment that has not been attempted in previous studies. In the proposed system, the user's participation in the platform can be realized through collective intelligence, and the opinions of the participating groups can be reflected in automated detection through deep learning. The deep-learning model guarantees effective profits from investing each other's resources through consensus and cooperation through blockchains in situations where they operate individually and suppresses situations where technologies are exposed.

This study is structured as follows. Section 2 describes the relevant techniques and research utilized in the study, and Section 3 provides detailed information on the structure, components, and operation method of the proposed system. Section 4 confirms the feasibility of this study through the implementation and experimentation of the proposed system, Section 5 presents the discussion and limitation of proposal, and Section 6 presents the conclusions.

2. Related Work

2.1. Blockchain

A blockchain is a type of distributed ledger technology that Satoshi Nakamoto first introduced in 2009 [6]. A block, which is a bundle of transactions, such as those occurring within the network, is submitted to the network along with its hash value and added to the blockchain, continuing the record. It is a technology that preserves its integrity by making the connected blocks mutually verifiable through the hash value. Furthermore, based on the P2P network, each node stores blocks with the same content to ensure permanence and transparency in transaction verification.

Blockchains can be divided into public networks and private networks. Blockchains that are widely used and anyone can participate in, such as Ethereum [7] and Bitcoin [8], belong to public networks and grant block generation authority through the Proof of Work (PoW) [9] consensus algorithm. The PoW consensus algorithm is a process of deriving a common agreement among unreliable participants on a blockchain network and electing a fair miner. Participants elect a miner through a competition to find a hash value that matches the nonce value and verify their participation in the blockchain through a GRAM-based hash operation, commonly referred to as GPU mining. In this process, public blockchains grant cryptocurrency as a reward to miners and induce competition, leading to correct consensus within the network. Unlike public blockchains, private networks are closed networks consisting only of designated participants. A network consisting of only designated participants may have different authorities, depending on their position, and a designated or delegated validator produces a block. Although private blockchains are distributed blockchains, there is an argument that they are not decentralized.

This study uses Hyperledger Fabric [10] for agreement between service providers. Hyperledger Fabric is a private blockchain platform based on open source that is most commonly used by enterprises as part of the blockchain technology development project of

the Linux Foundation for enterprises. Unlike Ethereum and Bitcoin, it is a permissioned network that uses certificates and PKI to verify participants. Figure 1 shows the rough structure of a Hyperledger Fabric network. Hyperledger Fabric manages participants through a membership service provider (MSP), an authentication management system in which the roles and authority of nodes in the network are defined. Unlike PoW-based public blockchains, it consumes significantly less load and resources than the PoW consensus algorithm and shows very fast speed by adopting a consensus method called ordering, where the orderer collects transactions from peers through the ordering service, creates blocks, and distributes them to the network. Network components are largely composed of clients, peers, and ordering service nodes (OSNs), each performing transaction generation, chaincode operation, and block generation through consensus. Although Hyperledger Fabric is a private network, it is a consortium blockchain with the characteristics of a consortium. The form of consortium blockchain is useful for complex organizational environments operating within the same industry and fields that require common transactions and information.

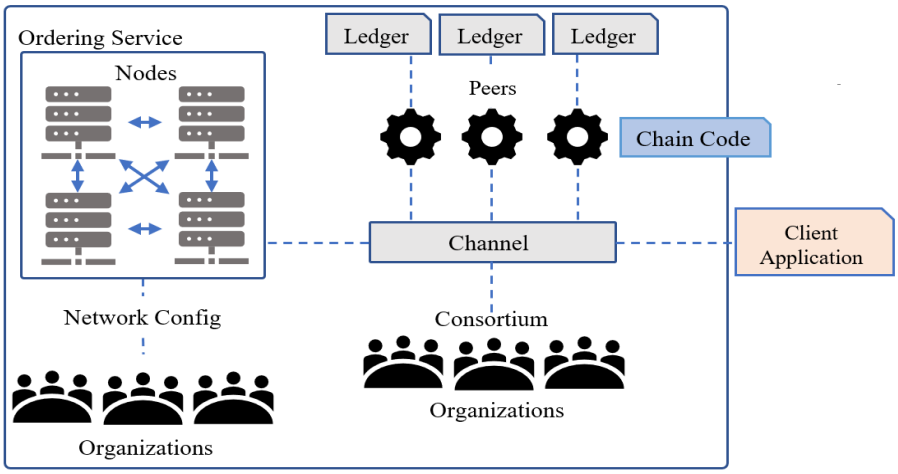

Figure 1. Configuration of a Hyperledger Fabric blockchain network.

Hyperledger Fabric supports the creation and deployment of a smart contract originating in Ethereum, which is called chaincode. The created chaincode is stored in the distributed ledger of the blockchain to ensure the integrity of the results of automated execution and input. Smart contracts operate decentralized applications in a blockchain environment, and code and execution environments are open to ensure reliability. Hyperledger Fabric processes transactions for chaincode operation in three steps (execution-ordering-validation) for the safe execution of chaincode. Recently, a study [11] by Sánchez-Gómez et al. for systematic consideration of design and testing of smart contracts was presented. This study suggests the improvement direction of code verification and verification in the development of smart contracts according to software development life cycle (SDLC). This study explores the smart contracts of several studies based on systematic literature review (SLR) and calls for the development of a design model of smart contracts in the DLT environment. Górski's study [12] shows that the verification and evaluation mechanism can be performed in a short time by proposing a method for actual testing of smart contracts. The above study shows the advantages of model-based design in performing smart contract tests using real cases as an example, and shows that verification and evaluation can be performed in one more test case than verification rules.

2.2. Collective Intelligence

To effectively utilize it in voting through the collective intelligence of the proposed system, the efficiency of the collective intelligence itself and the resulting efficacy were investigated. Modern society has entered an era of hyper-connection with the development

of the Internet network and the spread of mobile devices. Accordingly, research on collective intelligence, not just single human intelligence, is being conducted. Collective intelligence can be seen as collective participation, such as decision making through voting in modern society. Täuscher and Karl [13] studied the construction and utilization of collective intelligence through the crowd from the point of view of a company, presenting strategic goals as well as goals based on the empirical basis of the 80/20 Pareto principle [14] in the distribution of value creation among contributors. It also suggests an important point in the design of crowd-based business models (CBBM). Mann and Richard [15] presented a method of improving the accuracy of group prediction through optimal incentives when using collective intelligence.

Nguyen [16] proposed a blockchain-based trust model for information sharing in a crowd environment. This proposed a method for inferring reliable information in a collective intelligence environment and verifying its accuracy. Bonabeau [17] proposed that a loss of control was common to all forms of collective intelligence. The proposed system applies a reward system to prevent malicious users from occurring in the system through minimal control in a blockchain environment without administrators.

The integrity and permanence of the decision-making process of the user group are guaranteed through the blockchain, resulting in a contract environment where the results can be clearly disclosed to the service and users. This study implemented voting using collective intelligence in a blockchain environment. This is intended to demonstrate that the collective intelligence-based process operates normally in a distributed environment without a central manager and can make meaningful decisions. The model proposed in this study presents a method to increase the accuracy of collective prediction by giving users incentives through the cryptocurrency of a blockchain-based platform.

2.3. Blockchain-Based Reputation System

This study aims to supplement fairness through the integrated aggregation of deep learning and collective intelligence. For the efficient and effective operation of collective intelligence, the proposed system utilizes reputation-based voting weights and differentiates voting according to system contributions. This study refers to the most widely used reputation platforms for building reputation in the blockchain environment.

Steemit [18] is the most widely used blockchain-based reputation platform based on the Steem blockchain [19]. Steemit is a reputation-based copyright management platform, where reputation scores are generated by platform contribution activities, such as writing works and donations, and are used as personal trust in network participation. Steemit users directly own the copyright for the works they create on the platform and have the right to donate and curate their works. Through this process, it is possible to generate and manage profits for copyrighted works directly through the blockchain network without a central manager or service management entity. However, with Steemit, a copyright transfer contract is not possible, and problems may arise due to the absence of the self-purification function. The Steem network is a public blockchain based on delegated proof of stake (DPoS) [20], where network nodes to which users delegate mining rights create blocks through consensus. Stake, which is delegated by users in Steem's consensus algorithm, DPoS, refers to the contribution to the network, and this contribution refers to the delegation of participation in the consensus through cryptocurrency holdings.

Users of the Steem network evaluate and judge each other's work based on each other's reputation scores. The reputation score is judged as part of the contribution to the network, which is raised through methods such as writing, curating, and voting.

Figure 2 shows the distribution of reputation scores among Steem blockchain network users. In general, there are many users in the low-score range, which may be due to the method of calculating the reputation score from the raw-reputation data of the Steam network.

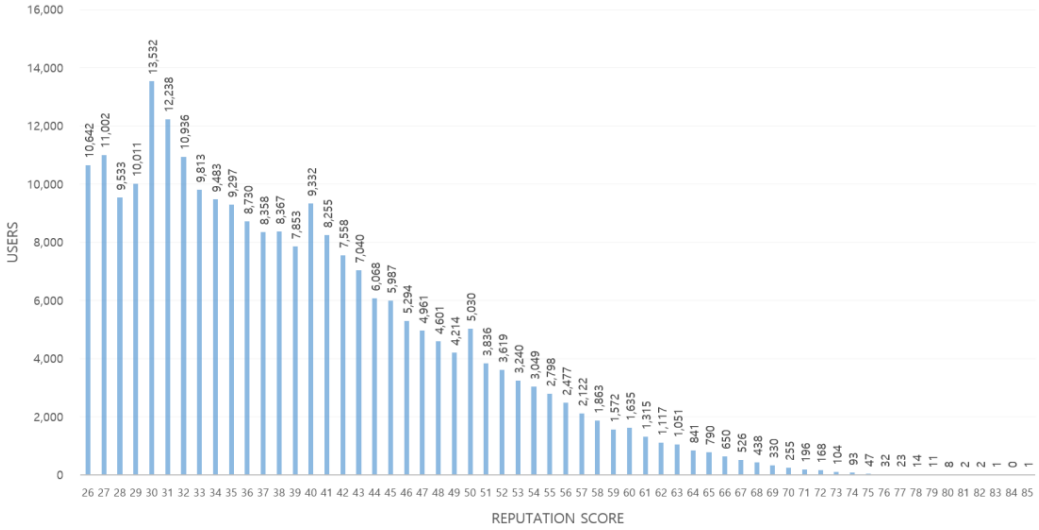

Figure 2. Distribution graph of user reputation scores in the Steem blockchain network.

This study utilizes reputation to give weight to user votes in user participation collective intelligence voting and borrows differential scores according to the user reputation distribution in the Steem network. The Steem reputation score is empirical reputation data generated by the free activities of blockchain users in actual services, which are subject to change as the volume of the proposed system expands. Steemit, a user community based on the Steam blockchain, implements a blockchain-based reputation system and uses the basic theory of collective intelligence. However, it is focused on creating and evaluating posts through user participation, and the evaluation itself does not change the platform. In this paper, we construct collective intelligence so that it can be established as a subject that operates and changes the system. In this process, the reputation system and the compensation system encourage collective intelligence on the blockchain to work properly.

2.4. Deepfake Detection

Deepfake is a video created by synthesizing a person's face or a specific part using artificial intelligence technology. Using deep learning, fake media are created by synthesizing the original video frame by frame with the face of another person. Deepfake mainly uses generative adversarial networks (GANs) [21]. Unnatural results may be obtained if the amount of training for the object to be synthesized is insufficient or if there are obstacles in the original video. However, with the recent development of artificial intelligence technology, the quality of synthesized videos has increased.

Following the development of technology for creating deepfakes, deepfake detection technologies are also being developed. Challenges are being faced through datasets such as DFDC [22] and FaceForensics++ [23]. Recently, studies using XceptionNet [24], Efficient ViT [25], and EfficientNet [26] have reported high detection performance. EfficientNet is a study that finds the optimal values of network depth, channel width, and input image resolution, which are factors that increase the size of the model in the recent research trend toward increasing the size of the model to improve performance through AutoML. SOTA has been achieved by proposing a compound scaling method that can efficiently create combinations. EfficientViT limits softmax attention, replaces it with linear attention, and improves local feature extraction capability through depthwise convolution. It also maintains global and local feature extraction capability while lowering linear computational complexity. XceptionNet uses the Inception V3 structure, modified from GoogLeNet's Inception Module. The 5×5 filter is transformed into a $3 \times 3 + 3 \times 3$ filter to effectively transform the parameters, and it distributes the two roles with cross-channel correlation while performing spatial correlation analysis of what a single convolution kernel does.

Through this, extreme inception (Xception), which claims that the mapping of cross-channel correlation and spatial correlation can be completely separated, is proposed.

The deep-learning model is evaluated for accuracy according to various scoring systems, including accuracy, recall, and precision. Most of the existing deep learning research for deepfake detection [22–26] has been conducted to improve accuracy, but the reliability of the deep-learning model itself has been poorly studied. It is difficult for general users to secure hardware to operate the deep-learning model, and there are budget and time limitations for operation. That is why users fully trust and use the detection and prediction models provided by businesses and services. We provide the reliability of the deep-learning model provided by companies through the proposed system as user-friendly information. In a blockchain environment, each company drives a deep-learning model as a node and strengthens the reliability of the deep-learning model itself through mutual node evaluation. In this study, it is used for an experiment to compare the accuracy of using the deep-learning model of the above study independently with the accuracy calculated by linking and integrating the deep-learning models and collective intelligence voting with the integrated voting through the proposed system. In the proposed system, each ICP drives a different deep-learning model for deepfake detection. Each deep-learning model has a reliability value through reliability evaluation, which is evaluated as a reliability contributing to the maintenance of the ICP system.

3. Deepfake Detection System through Collective Intelligence and Deep Learning

The goal of the proposed deepfake detection system (DDS) is to establish a fair and objective deepfake detection subject through collective intelligence and deep learning in a blockchain environment. Existing studies and services for detecting deepfakes relied on a single model in which the detection subject had to be completely trusted. As a result, existing studies have a problem in that they cannot respond when detection results are manipulated or errors occur. Through the proposed system, deep-learning models driven by each ICP side compete and sanction each other, complementing the transparency of the detection subject by integrating and using each result through a blockchain. This ensures the objectivity of detection results and establishes a healthy competition system among ICPs. The proposed system includes a service structure for efficient user participation and management of the blockchain network, a reputation model for discrimination in user voting and a voting weight management module through it, a node objectivity evaluation index representing the reliability according to the accuracy of the deep-learning model running on the ICP node, and the requirements of members to operate the proposal system based on a basic blockchain. It also includes a random voter selection and integration module for performing integrated voting. The proposed system is automatically operated without an administrator's response to the integrated voting request raised by random anonymous users, and it is executed through chaincode, a decentralized smart contract.

The integrated voting of the proposed system is calculated by the final aggregation of user voting based on collective intelligence and the result of the deep-learning model operated by ICP nodes, along with user reputation and a node objectivity evaluation index. User votes are selected based on RANDAO [27], a secure random number generation method in a blockchain environment, and are finally counted with vote weights based on user reputation. User reputation scores increase through system contribution activities, such as voting participation, vote suggestion, and deriving of significant vote, and are converged to a certain score after cumulative increase in the form of the exponential function.

The node model voting for the objectivity of integrated voting has the deep-learning model for deepfake detection running in each ICP node ("node model") as its voter. The detectors of the node model may be different, and the accuracy may also vary depending on the hardware specifications and detector specifications. The node model has an aggregate weight determined by the node objectivity evaluation index, which is an index that relatively evaluates whether its accuracy is sufficiently objective and accurate for integrated voting.

The proposed system was constructed in the form of a consortium blockchain network using Hyperledger Fabric. All functions of the system, except for the node model, are written in chaincode and run on the blockchain network. They are executed safely and flawlessly in a decentralized system, in which administrators cannot interfere. A detailed description of each function is provided in the corresponding section. We provide a list of abbreviations to help understand the terms used throughout the paper, and the list is shown in Abbreviations.

Blockchain-based software architectures are needed to prevent complex problems that can occur in distributed environments that are prone to systemic conflicts during operation at the design stage. Prior to designing our proposed system, we analyzed the design structure of the blockchain platform currently in service [28] and identified the structural characteristics of the Hyperledger Fabric. Through this, the chain code is designed and operated by reflecting the structural characteristics of the proposed system. In addition, the proposed system is constructed and tested by referring to the design process [28] for effectively designing a blockchain-based system at the system design stage.

3.1. Configuration of the Proposed System

Figure 3 shows the overall configuration of the proposed system. As shown in the figure, the proposed system is largely defined by three elements, which are defined as follows:

- Blockchain Network: A network composed of Hyperledger Fabric creates a consortium of nodes running on service providers. When a transaction generated by a network client is endorsed by an endorser and proposed, the transaction is sorted through a consensus algorithm and then generated as a block and propagated to peer nodes. Items, such as network rules, are determined through agreement between nodes.
- ICP Node: The node of the blockchain network of the proposed system can be any service, including video hosting services like YouTube and ICPs, involving all acts of posting, hosting, and downloading videos containing intellectual property as service content. In addition, the service node is assumed to have resources capable of driving a node model for deepfake video detection and deriving a result within a meaningful time. It is also assumed that the service will continuously evolve the node model according to the system's compensation policy. An ICP node is a peer node that is responsible for maintaining the network and executing chaincodes and has chaincodes installed to run on the network. It also maintains network security as an endorser by simulating the transactions generated.
- User: A user refers to the user of each ICP node constituting a blockchain network who can participate in the blockchain network as a client. A user may be a copyright holder who produces a work as a content creator. Users participating in the network as clients can participate as voters in the integrated vote for the deep-learning detection process.

3.2. New Content Registration Process for the Proposed System

The Hyperledger Fabric blockchain network has a transaction processing process through an ordering service, unlike the blockchain of the general PoW ecosystem. Figure 4 shows the process of a user registering new content in the proposed system and recording the information on the network based on Hyperledger Fabric.

When a user uploads a video through the ICP to which he or she belongs, the video data are stored in the database server of the ICP [29]. After ICP extracts and signs metadata from video data, it executes the registration chaincode for metadata registration as a client through nodes connected to the blockchain. The chaincode is installed on the peer node, and the chaincode execution transaction is transmitted to the endorser of the other peer node to receive endorsement of the safety of the chaincode before sending it to the orderer. The orderer generates blocks based on the consensus algorithm and sends them to peers. The peer's anchor receives the block sent from the orderer, confirms the block through the committer, and updates it on the blockchain. ICP extracts the metadata of these contents and registers them with the signature of the service on the blockchain ledger. The metadata

registration process is performed by the ICP as a single unit by collecting the content registered by the user of the corresponding service.

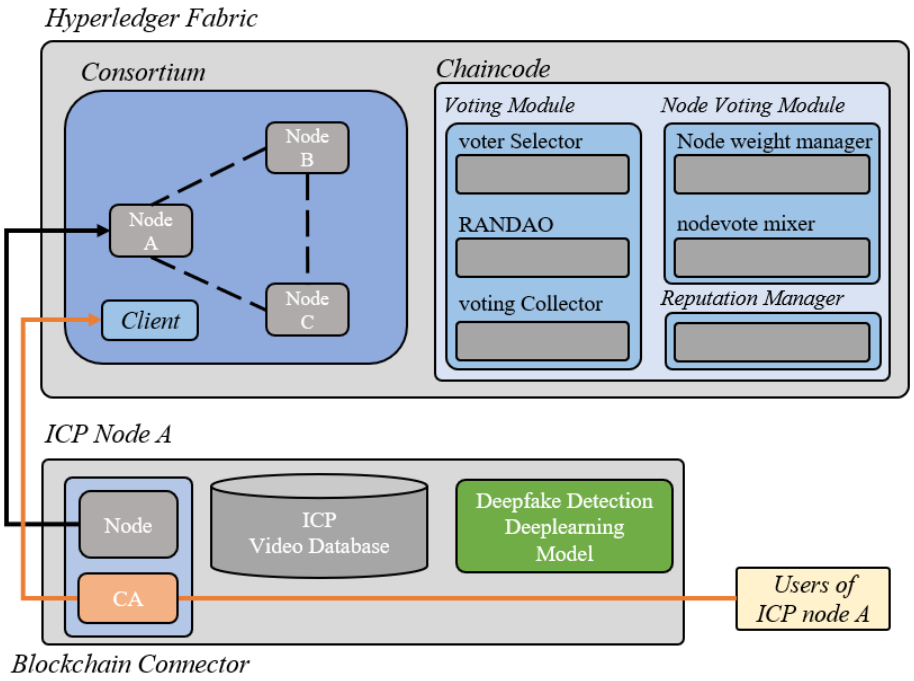

Figure 3. The structure of the proposed system.

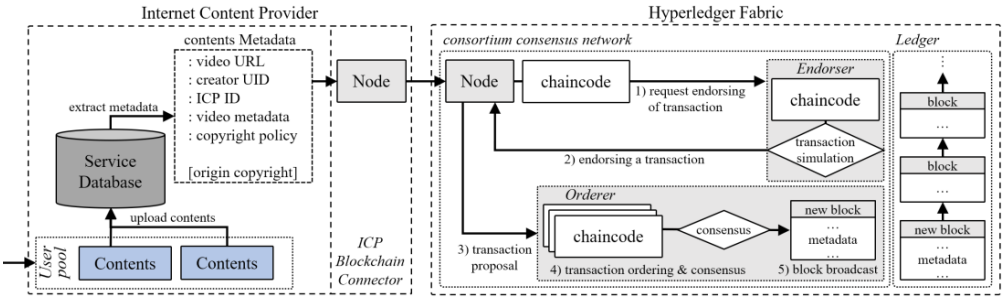

Figure 4. Transaction flow for registering the contents of ICP in the blockchain network in the proposed system and its chaincode.

3.3. Integrated Vote

The integrated vote in this study refers to the self-purification action of the system performed by the objection flag, which is a function of an objection raised by a user, a client, for suspected deepfake content registered in the system. Integrated voting proceeds through objections raised by unspecified users consuming ICP content that participate as nodes in the system registered through Section 3.2 above. According to the collective selfishness of crab mentality [30], content registered in ICP shows a proportionate frequency of censorship by users as awareness and profitability increase. Accordingly, content with high topicality has a high level of censorship criteria and a high possibility of user objection. The integrated voting process due to the user's objection is automatically performed on the network through the chain code and determines the authenticity of the deep fake video registered in the ICP to cause self-purification.

Integrated voting is performed by mixing user voting based on collective intelligence and node model voting based on node models operated independently by each ICP along with their respective weights at a certain ratio. The user vote part of the integrated vote for self-purification uses a module to randomly select unspecified users to operate fairly and safely in the blockchain environment In addition, voting weight is given to voters

based on their reputation, which is given according to the degree of contribution to the system to supplement voting reliability. Here, reputation is increased through content registration, voting participation, correct prediction of voting results, and objection to deepfake content, which are actions that contribute to the system for the survival of the blockchain-based network ecosystem. Node model voting based on deep-learning models powered by ICPs allows each ICP to independently inspect content suggested by users through different deep-learning models. ICP can select different deep-learning models, and only the results calculated from the deep-learning model are submitted in the form of a black box for the network parameters and structure of the deep-learning model. Since ICP does not disclose the technical specifications of deep-learning models, it is assumed that each deep-learning model will be developed competitively. The inspection results of each deep-learning model are added to the integrated vote through the node weight, which represents node reliability. Integrated voting for self-purification functions within the system integrates and aggregates the node models of users and ICPs at a certain rate. Here, the aggregation ratio is 70% clients and 30% node models. In addition to the 80/20 rule, this is a ratio determined to maintain the initial stability of the ecosystem and can be adjusted through consensus among blockchain network participants.

Figure 5 shows the process of integrated voting. When an objection flag, a content objection function, is received from the flagger, a user who claims to have a problem with a specific content, the process for user voting and node model voting for integrated voting is automatically performed in the proposal system. The results of user voting and node model voting conducted through the proposed system are integrated into the final results through aggregation.

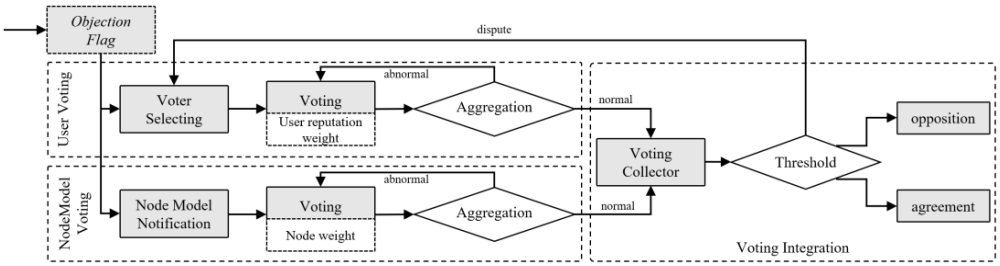

Figure 5. Integrated voting process of the proposed system.

Sections 3.3.1 and 3.3.2 describe user voting, in which users participate as clients in the integrated voting system, and node model voting, in which the node model, a deep-learning model driven by ICP, a network peer node, participates. The process of integrated counting through voter election and voting weight is explained.

3.3.1. Configuration of User Voting

Fair Election of Voters in a Blockchain Environment

Prior to conducting user voting based on collective intelligence for integrated voting, a way to fairly and safely elect voters in a blockchain environment is needed to ensure objectivity and neutrality in voting. In this study, random numbers are generated through RANDAO, a secure random number generation method in a blockchain environment, and voters are elected based on them.

Figure 6 shows the operation process of RANDAO, and Equation (1) shows the final random number calculation.

$$RN = S_a \bigoplus S_b \bigoplus S_c \quad (1)$$

RANDAO is a random number generation method in a blockchain environment, where the random number generators a, b, and c in the block n send the hash values of the random seeds S_a, S_b, and S_c of Secure Hash Algorithm 3 (SHA3) [31], SHA3(S_a), SHA3(S_b), and SHA3(S_c) to the blockchain. Then, after receiving the original seeds S_a, S_b, and S_c after k

blocks are promised in advance, the hash value is obtained, verified to match the previously recorded value, and the value obtained by Xoring the seed is used as the random number *RN*. The above method guarantees that a secure random number is generated if there are one or more flawless random number generators due to the impossibility of modifying the previous block and the integrity of the blockchain.

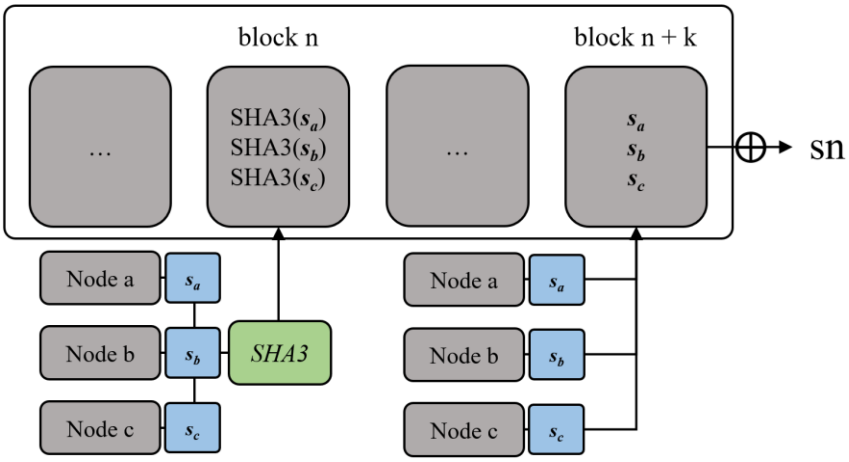

Figure 6. Random number generation through RANDAO in a blockchain environment.

The proposed system elects a voter by matching the generated random number with the system client UID using the peer node as a seed provider for random number generation. Algorithm 1 shows the pseudocode of voterSelector where the proposed system elects voters based on RANDAO.

Algorithm 1: voterSelector pseudo code

Input: *Nodes* = (n_1, n_2, n_3, ...) : ICP nodes
NV : Number of voters needed
UA = (u_1, u_2, u_3, ...) : Users Account Set
Output: *V* = (...) : Elected Voters

```
R ← RANDAO(Nodes)
for in NV do
    R ← SHA3(R)   #hashchain-style seed amplification with SHA3 Hashing
    targetFitting ← ABS(u1-R)
    for in k = 1 to n(UA) do   #approximation search for hashed R in UA
        gap ← ABS(uk-R)
        if (gap < targetFitting) then
            targetFitting ← gap
            targetVoter ← uk   #closest approximation of R
        end if
    end for
    V ← V ∪ targetVoter
end for
if (n(V) == NV) then
    return V
end if
```

The Allocation of Vote Weight for a Fair Vote

Voting based on collective intelligence using user participation requires a method to resist voting manipulation by malicious users. This study differentiates user voting weights

for voting stability. Through this, each user has a voting weight proportional to the system's contribution. This contribution is expressed as reputation, and accordingly, policies such as systematic differential compensation can be applied. According to game theory [32], users generally avoid actions that could be malicious to the system. The voting weight of the proposed system borrows from Steemit's reputation-based content management method. In order to calculate voting weight based on reputation, each user has their own reputation raw, which is calculated as reputation score through Equation (2).

$$RS_k = [log\{abs(RR_k)\} - 9,\ 0] \times \{(RR_k\ \geq\ 0)?\ 1 : -1\} \times 9 + 25 \tag{2}$$

Reputation raw is a value for gradually changing the reputation score in an integer range and is formed in the range of 10^9 or higher. Reputation raw is increased through content registration, voting participation, correct prediction of voting results, and objection to deepfake content, which are actions that contribute to the system. Through the formula, reputation raw is calculated as an integer reputation score in the range of 25 to 100. The equation shows the process of calculating the reputation score (RS) through the reputation raw (RR) of user k.

Figure 7 shows a graph of the reputation raw-reputation score in the Steem blockchain network that operates a reputation-based content management system through Equation (2), which is the reputation calculation formula of the proposed system. The graph shows the results of analyzing the reputation data of 242,361 Steemit community users recorded in the Steem network. The reputation raw-reputation score rises in proportion to the index, and the reputation raw in the high score range shows a very large difference.

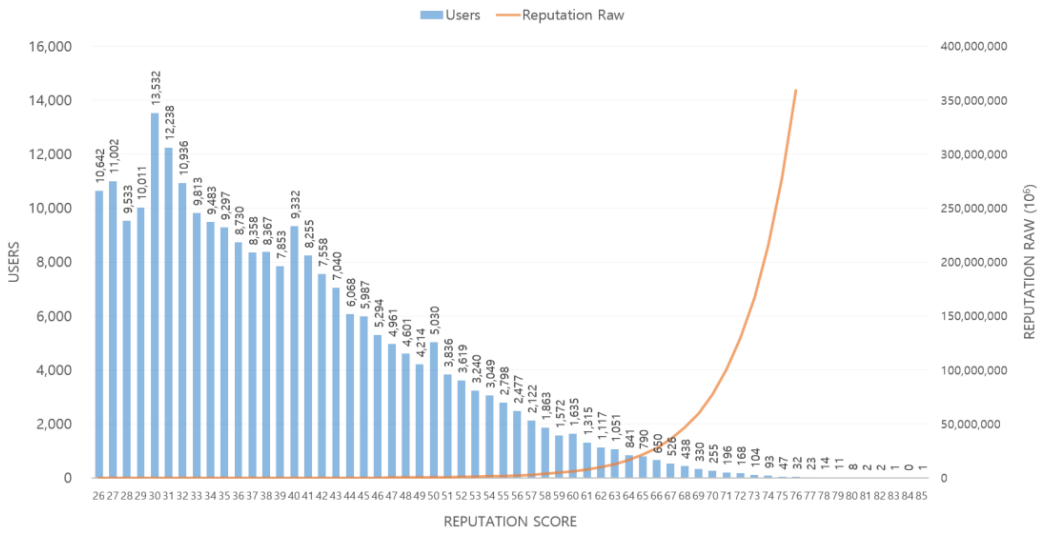

Figure 7. Reputation raw-reputation score graph.

The reputation score generated as shown in Figure 7 is calculated as a voting weight value in the range of 0.75 to 1.25, as shown in the reputation raw-voting weight graph of Figure 8 through Equation (3). Equation (2) is the reputation score-weight equation, where UW is the user weight, RS is the reputation score, and β is the conversion rate of RS-W, $\sqrt[RS-25]{5 \times 10^{-3}}$.

$$UW_k = (1.0965\beta)^{RS-25} + 0.745 \tag{3}$$

Through Equation (3), the voting weight has a gentler upward curve compared to the reputation raw graph. This alleviates the decrease in the number of people in the high score range of the reputation score due to the increase in the reputation raw value required to increase the reputation score. Through this, the distribution of voting weight is adjusted, and the phenomenon of biased voting results due to minority opinions is prevented. β, the

conversion rate of RS-W, represents the increase in weight proportional to the decrease in the number of people due to the increase in the reputation score.

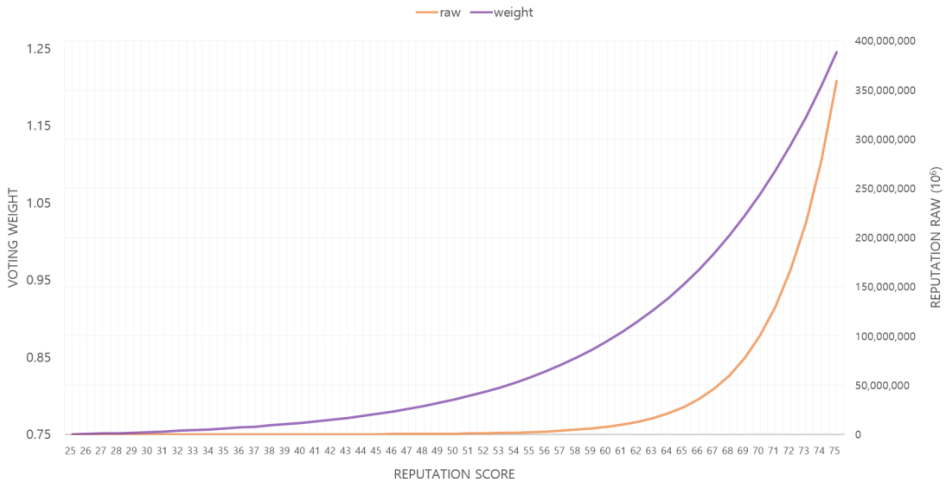

Figure 8. Reputation Score-Voting Weight graph.

Voting weight is increased through system contributions, and the effect on voting is adjusted according to the degree of user contribution. A higher reputation score means that more time has been spent on the system and more contributions have been made. Based on game theory, which assumes that the degree of malicious influence is low, the proportion of malicious votes in the voting results is reduced. This utilizes collective intelligence to prevent loss of control of lines that do not harm decentralization through the consortium blockchain and incentive system hosted by enterprise services, which is intended to enable users to act as consumers contributing to the existence of the platform.

3.3.2. Configuration of Node Model Voting

This section describes the node model voting of integrated voting for system self-purification functions. Node model voting has a 30% ratio, as mentioned above, in integrated voting, and aims to complement the neutrality and objectivity of integrated voting. Each node model undergoes node neutrality evaluation at regular intervals to maintain objectivity and accuracy and has a node weight accordingly.

Figure 9 shows the process configuration of node model voting. In the proposed system, different deep-learning models driven by ICP nodes can be used, and their accuracy can vary significantly depending on the hardware performance of the service server and deep-learning model's network parameters. The proposed system adjusts the influence of the node model on the overall integrated voting through neutrality and objectivity evaluation of node models of different specifications. The node objectivity evaluation index updates the voting weight through node evaluation by the possibility of continuous model development. Equation (4) shows the calculation of the voting weight through the node evaluation performed in the network.

$$NW[\alpha]_t = NW[\alpha]_{t-1} \times \left(V_{t-1}^{result} == NP[\alpha]_{t-1}^{result} \right) ? \; 1.01 \; : \; 1 \tag{4}$$

In Equation (4), NW means node weight, V means voting, and NP means node prediction, representing a situation in which the node model α progresses from the $t-1$ phase to the t phase. The weight of the node model shows a weight improvement of 1% point per successful prediction in each phase of the prediction reliability evaluation step through the equation. In phase $t-1$, when the final result of voting is the same as the predicted value of the node, the node weight of the corresponding node in phase t is increased. Nodevote weights can convene a committee composed of ICP nodes to re-evaluate the reliability of the node model when a specific weight is out of the normal

range. In the integrated voting stage, the weight is aggregated through the nodevote mixer along with the node model predictions and then integrated into the integrated voting along with the user collective intelligence voting.

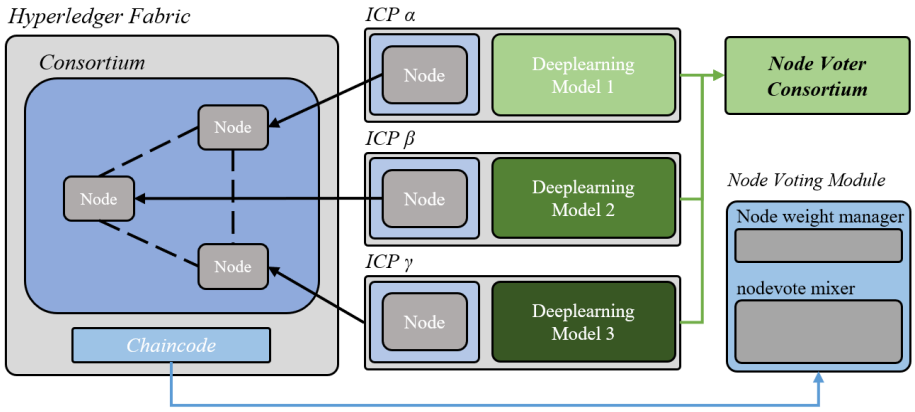

Figure 9. Node model voting process configuration.

Unlike the general ensemble of deep learning, the mixing of the nodevote mixer utilizes only the result of the deep-learning model in the form of a black box. The proposed system does not expose the strategy of each service by using only the results derived by the node model without limiting network parameters, such as loss of the node model. This induces a situation in which it is assumed that ICP nodes of hostile interest will continuously develop the node model for operating profit through a compensation policy in the blockchain environment. The weight of the node model can be evaluated as the technical reliability of the ICP, creating trust value for the enterprise. Nodevote mixing is described in detail in Section 3.3.3. Aggregation and Integrated Votes.

3.3.3. Composition of Integrated Vote

Progress of the Integrated Voting Process

Figure 10 shows the sequence of the entire process for integrated voting in the proposed deepfake detection system. The figure shows objections by users, election of voters, voting according to quarters, aggregation, and integration in the proposed system. The detailed process is as follows:

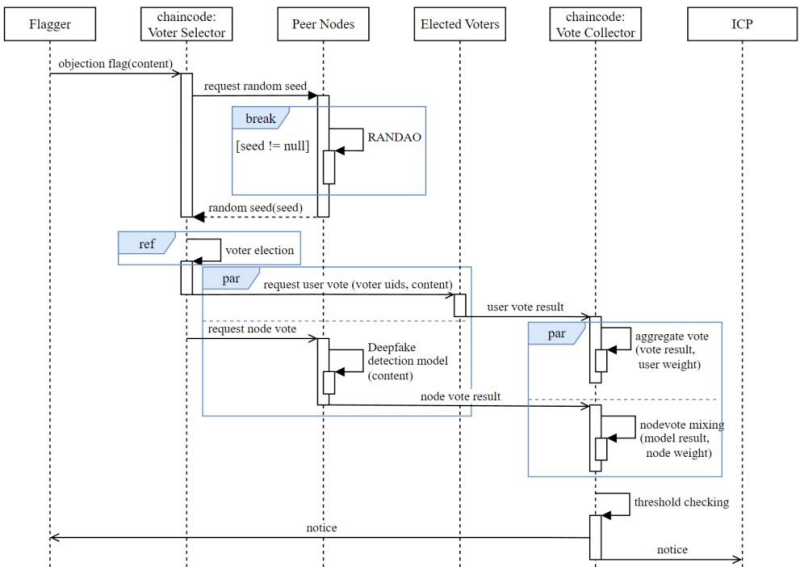

Figure 10. Process sequence diagram of the proposed system.

1. An objection flag, which is an objection to content suspected of being manipulated by an unspecified user of the network, is presented.
2. The network proceeds with the voterSelect process to initiate integrated voting.
3. The hash value of seed S for RANDAO is submitted from each ICP node by voterSelect.
4. After six blocks, the original seed *S* is submitted, and based on the random number *sn* generated by Xor, a voter is randomly selected by matching with the client UID.
5. Voters are notified, and voting takes place within 24 h. → User vote
6. Each ICP node is notified of the implementation of integrated voting and requests a decision value for the corresponding content. → Node model vote
7. The voting performed is integrated by adding voter weight and node weight through the voting collector.
8. Integration voting results for flags are judged as positive (+) or negative (−) values based on a specific threshold that exceeds the threshold.

Aggregation and Integration of Votes

Integrated voting, which is a self-purification function of the proposed system, is conducted by integrating the aggregate results of collective intelligence-based user voting and deep-learning-based node model voting on suspected deepfake contents. For user voting based on collective intelligence, after electing voters through the voter selector and requesting votes from client users in the network, they are aggregated together with the user weight in the voting collector and then integrated. Node model voting is based on deep-learning voting requests from ICP, a network peer node, and the node model running on the ICP node verifies suspicious content. The verification results of the node model are aggregated in the nodevote mixer along with the node weight of each node and then integrated with user votes in the voting collector.

Figure 11 shows the process of integrated voting in the proposed system. Integrated voting is initiated through the objection flag, which is an objection raised by the system user (client) for system self-purification. The flag transaction by the user is delivered to the system's main chaincode through the endorser and orderer, and the voting process is performed by branching into user voting and node model voting. The voting agenda is whether the target content has been modified by deepfake. The integrated voting of the proposed system is composed of *n* user voters and nodes, with *m* node models predicting the contents of *k* frames.

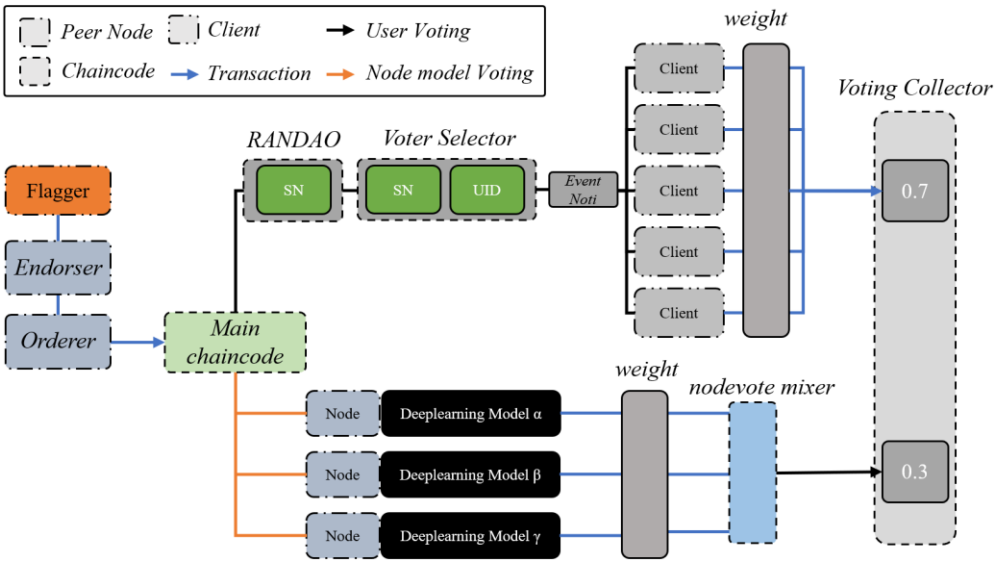

Figure 11. Integrated vote structure of the proposed deepfake detection system.

For user voting branched off from the voting process, after electing voters through the process of Section 3.3.1. Fair Election of Voters in Blockchain Environment, the event listener requests the elected voters to vote. The elected voters cast their votes during the time limit of the requested time. The number of voters elected and the voting time can be adjusted according to the size of the system. Equation (5) shows the user votes being counted in the voting collector.

$$VI^{voter} = \sum_{voter}^{n} (V[C]_i \times UW_i) \tag{5}$$

The voting integrated value of n voters VI^{voter} is calculated through the values of positive (1), abstention and invalidation (0), and negative (-1) of the voting result for the voting agenda content C and Equation (3). It is aggregated by adding the user weight (0.75–1.25) value. The results of user voting are counted in the voting collector and immediately reflected in the integrated vote.

The voting process requests node model voting from service nodes, which are peer nodes in the network, with branching through the system's main chaincode. The service node decides on the voting agenda for a limited time through a deep-learning model running in each locality. The node model is aggregated in the nodevote mixer along with the node weight generated through the node reliability evaluation in Equation (4). Equation (6) shows the aggregation of node model votes through the nodevode mixer.

$$VI^{node} = \sum_{node}^{m} \left\{ \frac{\sum_{f=0}^{k} F[C]_f[0,\ 0.1]}{k} \times NW[m] \right\} \tag{6}$$

Equation (6) shows the process of integrating votes for content composed of k frames with m node models. $F[C]_f$ refers to the prediction result of each node model and indicates whether there is a deep fake for each frame of content C submitted as a voting agenda. The prediction result of the node model uses the average result of all frames.

Node model votes aggregated in the nodevote mixer are merged with user votes through the voting collector. Equation (7) shows voting integration through voting collectors. In the integration stage, which is the end stage of integrated voting, the user voting value aggregated through user voting and the forecasting result of node model voting are integrated.

$$V^{value} = VI^{voter} + VI^{node} = \begin{cases} opposition, & V^{value} < 27 \\ dispute, & 27 \leq V^{value} \leq 73 \\ agreement, & 73 < V^{value} \end{cases} \tag{7}$$

The integration result, V^{value}, represents positive or negative votes to the voting agenda according to the threshold of 50 ± 23, and results below the threshold give a hold status to the voting agenda contents. The pending content is notified of each service so that the re-voting process can be performed by the service.

Compensation Policy

Table 1 shows the compensation policy according to participation in integrated voting for network users (clients). Incentives are provided according to the compensation policy to improve the accuracy of collective intelligence through the categorization of participation in integrated voting for the system self-purification of client users and compensation. A client who participates in integrated voting, a system-friendly act, and submits a vote in the same direction as the voting result can obtain an increase in reputation raw value and a self-purificator (SP) token, a cryptocurrency within Hyperledger Fabric, as a reward. Rewards are given to participants who voted that there would be no problem when there is no problem with the content and to the participant who voted that there would be a problem when there is indeed a problem with the content and the objector.

Table 1. Compensation policy according to the result of the integrated vote.

	Voting Result	
	Positive	**Negative**
Reward recipient	Flagger Positive Voter *	Negative Voter

* Voters who voted that the disputed content was indeed problematic.

Reputation raw, which is the base value of a user's reputation score, changes through system contributions. Here, system contribution behavior is defined as follows:

- Registering the contents in each ICP and registering the content information in the system through the ICP node.
- Participating in integrated voting-user voting to ensure smooth operation of the proposed system.
- Prediction accepted in the compensation range according to Table 1.
- Appeal and approval of suspected deepfake content.

4. Experiment and Results

4.1. Experiment Environment and Dataset

For the experiment and evaluation of the system proposed in this paper, implementation and experiments were carried out, and the environment is shown in Table 2. This study previously explained the system with n voters and m node models. In this section, an environment was configured with 100 voters and three blockchain peer nodes to implement and experiment with the proposed system.

Table 2. Experiment and implementation environment.

Blockchain	
Network Environment	Virtual Machine Docker Ubuntu 20.04 LTS Hyperledger Fabric 2.1.0 *
Chaincode Language	Golang
Node Model	
Machine Environment	i9-10980XE, RTX 3090 Anaconda Ubuntu 18.04.5 LTS cuda 11.1.1, pytorch 1.8.0
Language	python3

* https://github.com/hyperledger/fabric/releases/tag/v2.1.0 (accessed on 5 January 2023).

In this study, the experiment through the node model consisted of three peer nodes. The pretrained deep-learning models of α, β, and γ in Table 3 were assigned to each node model. The official AUC of the assigned deep-learning model was 0.919, 0.951, and 0.944, but an experiment was conducted to confirm the difference in AUC due to variations, such as differences in hardware.

The dataset for the experiment was composed, as shown in Table 4. The dataset was sliced into units of 1000 and composed of feedback epochs to check the real-time nature of node reliability evaluation. Feedback epoch refers to a unit of performing one feedback 1000 times, which refers to a series of processes including proposing voting, voting, aggregation, integration, and node reliability reflection, considering the integrated voting process for contents. This study assumes a node model that can use different deep-learning models. Since each model may have a difference in accuracy for the dataset, each model can derive fair results by using the datasets in Table 4.

Table 3. Deepfake detection model of ICP node for experiment.

Node	Node Model	Official AUC
α	Efficient Vision Transformer [25] [1]	0.919
β	Cross Efficient Vision Transformer [25] [2]	0.951
γ	Ensemble EfficientNet B4/B4ST/B4Att/B4AttST [26] [3]	0.944

[1] https://github.com/davide-coccomini/Combining-EfficientNet-and-Vision-Transformers-for-Video-Deepfake-Detection/tree/main/efficient-vit/ (accessed on 5 January 2023). [2] https://github.com/davide-coccomini/Combining-EfficientNet-and-Vision-Transformers-for-Video-Deepfake-Detection/tree/main/cross-efficient-vit/ (accessed on 5 January 2023). [3] https://github.com/polimi-ispl/icpr2020dfdc/ (accessed on 5 January 2023).

Table 4. Deepfake dataset for the experiment.

Num	Name	Volume
1	DFDC [22] [1]	124 k videos
2	Faceforensics++ [23] [2]	3000 videos

[1] https://ai.facebook.com/datasets/dfdc/ (accessed on 5 January 2023). [2] https://github.com/ondyari/FaceForensics/ (accessed on 5 January 2023).

4.2. Experiment

The experiments in this study were conducted under environmental assumptions. The assumptions were as follows. The network peer node, the ICP node, submits the *seed* to the network within normal time for smooth operation of RANDAO for voter election. In addition, the node model calculates the prediction result for the agenda content within a normal time limit and submits it to the blockchain network. In addition, the node model calculates the prediction result for the agenda content within a normal time limit of 1440 min and submits it to the blockchain network. In the experiment, the video of the dataset was divided into frames and used for each feedback epoch. In the experiment, the voting accuracy of user voting by system clients was assumed to be 92.5 ± 2.5%p. The experimental environment consisted of consensus through an ordering service, 100 voters, three peer nodes, and a node model in a Hyperledger Fabric blockchain network environment.

Table 5 shows a comparison between the area under the receiver operating characteristic (AUROC) of each deep-learning model and the AUC through the integrated voting process of the proposed system. Each node model was tested through a combined dataset and environment for the experiment in this study. In the experiment, each node sequentially occupied hardware resources for detection and decision making.

Table 5. AUC comparison between models for deepfake detection.

	ICP α	ICP β	ICP γ	ICP + Vote
Model	efficient ViT	Cross efficient ViT	efficientNetB4	our
AUC	0.8372	0.8809	0.9012	0.9263

Figure 12 shows the ROC curve and confusion matrix of the integrated voting of the proposed system. Through the ROC curve and confusion matrix, high accuracy is shown in the final decision through integrated voting when it is an actual deep fake video.

In the proposed system, when the average accuracy of the node model was 87% or higher, VI^{node} of V^{value} in Equation (7) for voting collection was 26.1044. Each node model is assumed to be given a node weight proportional to its accuracy by the reliability evaluation. The minimum value for exceeding the threshold-73 of VI^{voter} for the agreement of integrated voting is 46.8957. It is resistant to up to 16 malicious voters based on 100 voters in an experimental environment, assuming a voting accuracy of 92.5 ± 2.5%p.

Figure 13 shows the AUC change of the independent deep-learning model and the proposed system during each feedback epoch. The above figure shows the AUC change according to the data set of the final aggregation model of integrated voting and when the deep-learning models α, β, and γ of each node constituting integrated voting are

independently driven. The final aggregation value is a sum of the user vote aggregation result of Equation (5) and the nodevote mixing result of Equation (6). The user vote result value VI^{user} has a maximum value of 70, and the nodevote result value VI^{node} has a maximum value range of 30. The graph shows that the proposed model shows high and stable predictions on average. This shows that stable predictions can be made by integrating user votes and the results of each node model according to node reliability. Since the video contents registered in the ICP are unpredictable and a very large amount of data is uploaded, those with average stable predictions and small changes can have a high safety level for incident response. The proposed detection process shows an AUC in the range of 91–95 in response to changes in the data set through experiments. Through this, it is confirmed that the detection process of the proposed system performs objective and neutral stable prediction because the vibration width of the AUC according to the data fluctuation is smaller than that of the individual deep-learning model.

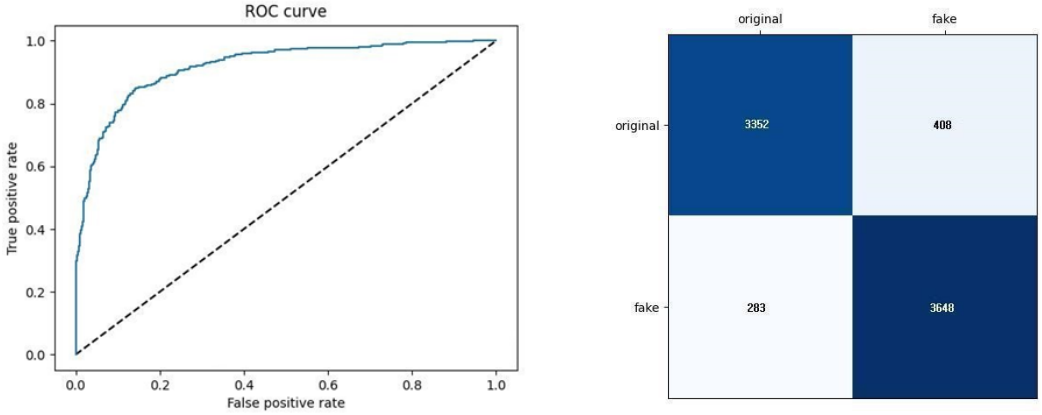

Figure 12. (**a**) ROC curve of ICP node + vote and (**b**) confusion matrix of integrated voting of the proposed system.

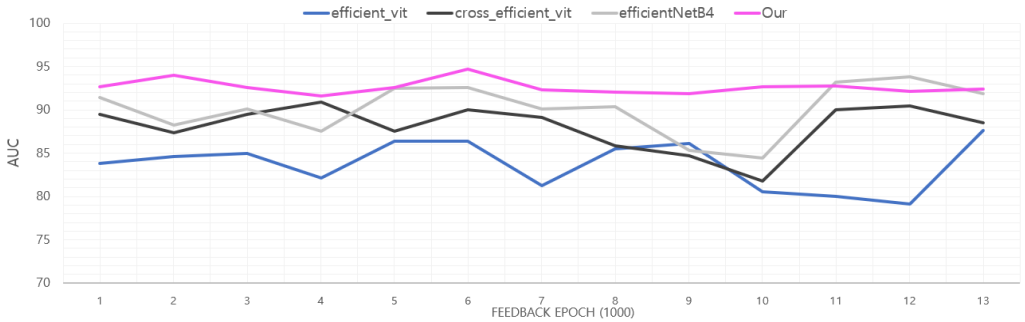

Figure 13. Graph of AUC change according to feedback epoch in integrated voting.

Figure 14 shows the change in the node weight value of the node model during each feedback epoch for the node reliability evaluation. As the experiment proceeds with the node model fixed, the node weight converges similarly to the accuracy distribution of each model. EfficientViT, cross-EfficientViT, and efficientNetB4 were assigned as the three nodes in the experimental environment. The hardware and environmental performance of each node were identical. Figure 14 shows that the real-time node reliability evaluation for node weight measurement is performed stably as the epoch progresses. The accuracy of each node model is directly related to node reliability, correcting the node model's contribution to the integrated vote. Each feedback of the node reliability evaluation has 1000 frames of examples and a 1%p change range.

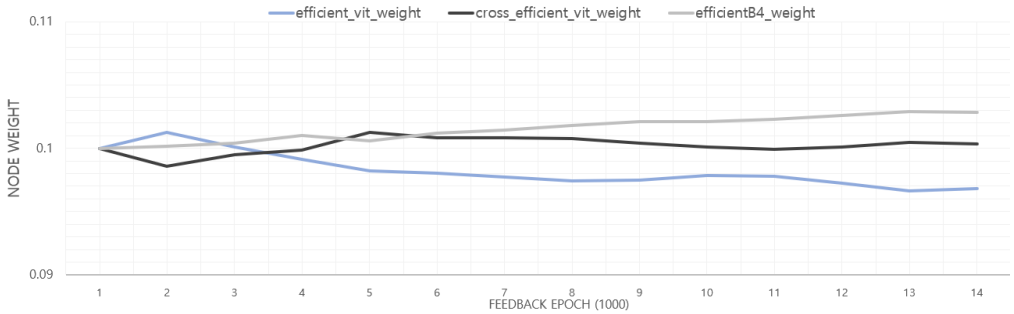

Figure 14. Graph of weight change of node model according to feedback epoch.

5. Discussions and Limitations

In this paper, we propose a system for detecting deep fake content in an environment where ICP services are provided to users. The proposed system is constructed through a blockchain, and the final prediction is performed by integrating collective intelligence-based voting and the prediction results through the deep-learning model. We considered how to use collective intelligence-based voting significantly throughout the paper. This paper attempts to solve the problem of individual information (manipulation of information, private use of results, possibility of incitement) and independence (dependence on other people's opinions, bias of results by minority opinions) [33] that occur in voting through collective intelligence. Each problem was solved through the integration of the prediction results of the deep-learning model by the machine, the voting process, and contribution weighting in the blockchain environment. In addition, by implementing collective intelligence in a blockchain environment, it transparently solves the distributed environment (the way to solve the problem should not be concentrated in one place) and the economic cohesion problem that can occur in incentive-based collective intelligence optimization. From an environmental point of view, the blockchain allows collective intelligence and deep learning, which are the methods that make up the system, to work complementarily, complementing its reliability and transparency. This solves the social costs caused by the non-trust relationship between the existing companies and the absence of the cooperation system. As a result, agreements are drawn based on meaningful results and transparency of the process at low levels of mutual trust, and complementary partnerships are established. Deepfake detection through deep learning uses detection and prediction results from individual detection subjects when used in terms of enterprise and service. The proposed system solves the problem of the total trust of the detection subject that occurs in this dependent detection environment. By designing a blockchain-based ensemble network based on a deep-learning model configured as an independent node as a peer, we overcome the problem of deep learning by utilizing the blockchain in a way that has not been attempted before. We anticipate that these attempts may be in line with the web3.0 framework where users participate in blockchain and manage data independently and explore new possibilities for blockchain.

We have limitations in constructing a collective intelligence experimental environment in implementing and experimenting with the proposed system. Collective intelligence can be influenced by the background of the times and social trends, and it can change according to the individual characteristics of the participating group. Therefore, unlike implementing collective intelligence, experimentation needs to be developed through actual users and platforms. In order to construct the experimental environment, we assume voting figures by referring to the percentage of user opinions that are not malicious to the system in wiki and crowdsourcing, which are collective intelligence-based platforms. However, there is a limit to reflecting the characteristics of collective intelligence to determine whether a problem is right or wrong in the proposed system. In future research, we will collect long-term data for the reliability of collective intelligence to improve the system. Data will be collected in various ways by adjusting conflicting opinion rates within the participating group and by variables such as the age and gender of the group. Through this, this study intends to

supplement the basis of the user voting participation rate and the positive (+) direction voting success rate presented in this paper. In addition, the proposed system experiments with the deep-learning model of the ICP node, which is a peer node, in a fixed model state using a major deep-learning model known to date. Deep-learning models and processes for deep fake detection will be further developed in the future, and we will apply an improved deep-learning model to the node model accordingly. Through this, the detection process prediction accuracy of the proposed system will show improved prediction accuracy in proportion to the accuracy of individual models, and we will build a more improved system by adjusting the node weight and composition ratio through experiments.

6. Conclusions

This study proposed a system to prevent social confusion caused by manipulated video content in the information hyper-connected era through the development of ICT technology and the Fourth Industrial Revolution. The proposed system detects deepfake videos shared on the ICP service and predicts whether they were deepfakes. Furthermore, it presents a solution to the reliability and fairness of the detection subject, which is a problem of various studies on deep fake detection that have been studied in the past.

The proposed system uses a method for detecting deepfake images by ensembling the detection results of collective intelligence and deep-learning models in a blockchain environment. In the proposed method, the blockchain network uses independent ICPs as nodes that drive each deep-learning model for deepfake detection. Through this, it realizes the value sharing of censorship and image information management functions that ICPs with conflicting interests keep to the inside of the service. ICP users participate as voters in user voting based on collective intelligence to supplement the fairness and universality of detection results. Each detection result aggregated through user voting and node model voting is integrated into the blockchain to ensure the reliability and transparency of the final decision. It operates on the basis of smart contracts on the blockchain and guarantees the integrity of the process.

As a result of implementing and testing the proposed system in this study, it was confirmed that the method of integrating the detection results of the deep-learning model of the proposed system by counting and integrating user votes showed higher accuracy than the existing individual and independent deepfake detection models. The detection process performed by the system is automated through a blockchain network, which is not controlled by external entities or administrators without authorization, and all users can check the execution process. This ensures the transparency and reliability of the detection process, and the reliability of the detection subject is secured through the voting process. When the three deep-learning models used for the experiment, efficientViT, cross-efficientViT, and efficientNetB4, show AUC of 0.8372, 0.8809, and 0.9012 in the same hardware environment and data environment, the results obtained through the integrated voting of the proposed system are 0.9263. In addition, it was confirmed that the detection process has objectivity and high reliability by showing stable prediction accuracy in the range of 91–95, unlike the individually driven model in the experiment, according to the change of the data group.

Through the proposed system, the possibility of organic cooperation between blockchain and artificial intelligence was confirmed, suggesting that the reliability and integrity of artificial intelligence and the effective guarantee of the permanence of version control were possible through blockchain. In addition, by using blockchains, a consortium for a common purpose between companies was created, allowing hostile companies to create common values. In addition, it was possible to build an environment for the effective use of collective intelligence through a blockchain network. Although the proposed system is focused on detecting deepfakes, it is expected to be helpful when applied to decision-making systems in various domains.

We will supplement the experimental evidence through the practical development of collective intelligence to overcome the limitations of the experimental environment in

future research progress. In addition, beyond aggregating the results of each deep-learning model in black-box form, we will design a large-scale federated learning environment for super-large artificial intelligence by forming a consortium for each machine to collect, learn, and detect data through the blockchain.

Author Contributions: Conceptualization, H.K.; Methodology, N.C.; software, N.C.; validation, N.C. and H.K.; writing—original draft, N.C. and H.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Kyonggi University Research Grant 2020.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

DDS	Deepfake Detection System
ICP	Internet Content Provider
SHA3	Secure Hash Algorithm 3
RR	Reputation Raw
RS	Reputation Score
UW	User Weight
NW	Node Weight
NP	Node Prediction
V	Voting
VI	Voting Integrated value

References

1. The Impact of the Internet on Society: A Global Perspective. OpenMind. Available online: https://www.bbvaopenmind.com/en/articles/the-impact-of-the-internet-on-society-a-global-perspective/ (accessed on 5 January 2023).
2. Repez, C.P.F.; Popescu, M.-M. Social Media and the Threats against Human Security—Deepfake and Fake News. In *Romanian Military Thinking International Scientific Conference Proceedings*; Military Strategy Coordinates under the Circumstances of a Synergistic Approach to Resilience in the Security Field; Centrul Tehnic-Editorial al Armatei; Ministry of Defense: București, Romania, 2020; pp. 44–55. ISSN 2668-8115.
3. Li, Y.; Zhang, C.; Sun, P.; Ke, L.; Ju, Y.; Qi, H.; Lyu, S. DeepFake-o-Meter: An Open Platform for DeepFake Detection. In Proceedings of the 2021 IEEE Security and Privacy Workshops (SPW), Virtual, 27 May 2021; pp. 277–281. [CrossRef]
4. Hasan, H.R.; Salah, K. Combating Deepfake Videos Using Blockchain and Smart Contracts. *IEEE Access* **2019**, *7*, 41596–41606. [CrossRef]
5. Patil, U.; Chouragade, P.M. Blockchain Based Approach for Tackling Deepfake Videos. *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.* **2021**, *7*, 342–347. [CrossRef]
6. Sunyaev, A. Distributed Ledger Technology. In *Internet Computing*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 265–299. [CrossRef]
7. What Is Ethereum? *Ethereum.Org*. Available online: https://ethereum.org/en/what-is-ethereum/ (accessed on 5 January 2023).
8. Nakamoto, S. Bitcoin: A Peer-to-Peer Electronic Cash System. *Decentralized Bus. Rev.* **2008**, 21260. Available online: https://bitcoin.org/bitcoin.pdf (accessed on 5 January 2023).
9. Gervais, A.; Karame, G.O.; Wüst, K.; Glykantzis, V.; Ritzdorf, H.; Capkun, S. On the Security and Performance of Proof of Work Blockchains. In Proceedings of the 2016 ACM SIGSAC conference on computer and communications security, Vienna, Austria, 24–28 October 2016; pp. 3–16. [CrossRef]
10. Hyperledger Fabric – Hyperledger Foundation. Available online: https://www.hyperledger.org/use/fabric (accessed on 5 January 2023).
11. Sanchez-Gomez, N.; Torres-Valderrama, J.; Garcia-Garcia, J.A.; Gutierrez, J.J.; Escalona, M.J. Model-Based Software Design and Testing in Blockchain Smart Contracts: A Systematic Literature Review. *IEEE Access* **2020**, *8*, 164556–164569. [CrossRef]
12. Górski, T. The k + 1 Symmetric Test Pattern for Smart Contracts. *Symmetry* **2022**, *14*, 1686. [CrossRef]
13. Täuscher, K. Leveraging Collective Intelligence: How to Design and Manage Crowd-Based Business Models. *Bus. Horiz.* **2017**, *60*, 237–245. [CrossRef]
14. Dunford, R.; Su, Q.; Tamang, E. *The Pareto Principle*; PEARL: Plymouth, UK, 2014; ISSN 1754-2383.
15. Mann, R.P.; Helbing, D. Optimal Incentives for Collective Intelligence. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 5077–5082. [CrossRef] [PubMed]

16. Nguyen, M.; Bai, Q.; Yu, J. A Blockchain-Based Trust Model for Crowd Environments. In Proceedings of the Australasian Computer Science Week Multiconference, Melbourne, Australia, 4–6 February 2020; pp. 1–7. [CrossRef]
17. Bonabeau, E. Decisions 2.0: The Power of Collective Intelligence. *MIT Sloan Manag. Rev.* **2009**, *50*, 45.
18. Trending Posts—Steemit. Available online: https://steemit.com/ (accessed on 5 January 2023).
19. Powering Communities and Opportunities—Steem. Available online: https://steem.com/ (accessed on 5 January 2023).
20. Snider, M.; Samani, K.; Jain, T. Delegated Proof of Stake: Features & Tradeoffs. Multicoin Capital: New York, NY, USA, 2018; Available online: https://multicoin.capital/2018/03/02/delegated-proof-stake-features-tradeoffs/ (accessed on 5 January 2023).
21. Creswell, A.; White, T.; Dumoulin, V.; Arulkumaran, K.; Sengupta, B.; Bharath, A.A. Generative Adversarial Networks: An Overview. *IEEE Signal Process. Mag.* **2018**, *35*, 53–65. [CrossRef]
22. DeepFake Detection Challenge Dataset. Available online: https://ai.facebook.com/datasets/dfdc/ (accessed on 5 January 2023).
23. GitHub - Ondyari/FaceForensics: Github of the FaceForensics Dataset. Available online: https://github.com/ondyari/FaceForensics (accessed on 5 January 2023).
24. Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258. [CrossRef]
25. Coccomini, D.A.; Messina, N.; Gennaro, C.; Falchi, F. Combining Efficientnet and Vision Transformers for Video Deepfake Detection. In *Proceedings of the International Conference on Image Analysis and Processing*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 219–229. [CrossRef]
26. Bonettini, N.; Cannas, E.D.; Mandelli, S.; Bondi, L.; Bestagini, P.; Tubaro, S. Video Face Manipulation Detection through Ensemble of Cnns. In Proceedings of the 2020 25th international conference on pattern recognition (ICPR), Milan, Italy, 10–15 January 2021; pp. 5012–5019. [CrossRef]
27. Blockchain Based Verifiable Random Number Generator. Available online: https://www.randao.org/ (accessed on 5 January 2023).
28. Xu, X.; Weber, I.; Staples, M. Existing Blockchain Platforms. In *Architecture for Blockchain Applications*; Springer International Publishing: Cham, Switzerland, 2019; pp. 27–44. ISBN 978-3-030-03034-6.
29. Choi, N. Blockchain-Based Over-the-Service Copyright Protection and Management System. *J. Korean Inst. Inf. Technol.* **2021**, *19*, 123–132. [CrossRef]
30. Ltd, I.-I.B. The "Crabs in a Bucket" Mentality in Healthcare Personnel: A Phenomenological Study. *Hitit Üniversitesi Sosyal Bilimler Enstitüsü Dergisi* **2019**, *12*, 618–630. [CrossRef]
31. Dworkin, M.J. SHA-3 Standard: Permutation-Based Hash and Extendable-Output Functions. *NIST* **2015**, FIPS-202. [CrossRef]
32. Owen, G. *Game Theory*; Emerald Group Publishing: Bingley, UK, 2013; ISSN 1-78190-508-8.
33. Centola, D. The Network Science of Collective Intelligence. *Trends Cogn. Sci.* **2022**, *26*, 923–941. [CrossRef]