SinkrOn
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15373
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Integration of Machine Learning and Blockchain for Forest Fire Risk Prediction

Nursetiaji Ramadhani[1], Ira Diana Sholihati[2], Djarot Hindarto[3]*, Asrul Sani[4]
[1,2,3]Informatika, Fakultas Teknologi Komunikasi dan Informatika, Universitas Nasional
[4]Magister Teknologi Informasi, Fakultas Teknologi Komunikasi dan Informatika, Universitas Nasional
[1]nursetiajiramadhani2022@student.unas.ac.id, [2]iradiana2803@gmail.com, [3]djarot.hindarto@civitas.unas.ac.id,
[4]asrul.sani@civitas.unas.ac.id

Submitted :Oct 16, 2025 | **Accepted** : Oct 23, 2025 | **Published** : Nov 11, 2025

Abstract: This study presents an integrated framework combining machine learning and blockchain technology to enhance the accuracy, transparency, and reliability of forest fire risk prediction in tropical regions. Using geospatial and climatological datasets from Google Earth Engine (GEE), two ensemble algorithms—Random Forest (RF) and Extreme Gradient Boosting (XGBoost)—were trained to model spatial fire susceptibility based on variables such as temperature, humidity, rainfall, wind speed, and vegetation index (NDVI). The RF model effectively identified low-risk areas but was less sensitive to minority high-risk classes, while XGBoost demonstrated superior adaptability in handling class imbalance and achieved more balanced performance across all categories. To ensure data authenticity and traceability, the prediction results were validated and recorded on the Ethereum blockchain using smart contracts. Each prediction output was transformed into a cryptographic hash (SHA-256) to guarantee immutability and verifiability. The integration of machine learning with blockchain establishes a decentralized, tamper-proof, and verifiable prediction system, promoting data integrity and transparency in environmental monitoring. Overall, this research introduces a novel "verifiable prediction pipeline" that advances both artificial intelligence and blockchain applications in environmental informatics, supporting proactive and accountable forest fire mitigation strategies.

Keywords: Forest Fire Prediction; Machine Learning; Random Forest; XGBoost; Blockchain; Smart Contract; Geospatial Data; Data Transparency

INTRODUCTION

Forest fires are among the most critical environmental challenges in tropical regions, particularly in Indonesia, where vast and biodiverse forest ecosystems are highly vulnerable to climate change and human activity. These fires cause significant ecological degradation and have major socioeconomic and health consequences. According to the National Disaster Management Agency (Badan Nasional Penanggulangan Bencana, 2024), more than 300 forest and land fires were recorded across Indonesia in 2024, resulting in ecosystem damage, economic losses, and health problems related to haze and air pollution. The increasing frequency of droughts linked to the El Niño phenomenon, combined with unsustainable land clearing and weak spatial monitoring, has worsened fire occurrence in recent years. Therefore, accurately and promptly predicting forest fire risk is vital for effective forest management and disaster mitigation in tropical regions.

In recent years, machine learning (ML) has emerged as a powerful tool for environmental modeling and fire risk prediction, as it can capture nonlinear relationships among diverse environmental variables. Two of the most widely used and effective algorithms are Random Forest (RF) (Firmansyah et al., 2024) and Extreme Gradient Boosting (XGBoost) (Karurung et al., 2025). Both algorithms demonstrate high accuracy in mapping fire susceptibility using geospatial indicators such as Normalized Difference Vegetation Index (NDVI), Land Surface Temperature (LST), rainfall, soil moisture, and proximity to human activity (Zou et al., 2023), (Karurung et al., 2025). These models effectively represent complex interactions between climate, vegetation, and topography to produce spatially explicit predictions. However, despite their reliability, most predictive systems still rely on centralized data architectures, leading to issues of data security, authenticity, and transparency key concerns for multi-agency decision-making in environmental management.

Previous studies have primarily focused on increasing predictive accuracy through hyperparameter tuning, feature selection, or ensemble learning techniques. Yet few have considered the integrity and traceability of

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15373

e-ISSN : 2541-2019
p-ISSN : 2541-044X

predictive data. As environmental prediction increasingly depends on cloud-based infrastructures, risks of data manipulation, loss, and limited accountability have grown. Centralized systems lack sufficient auditing mechanisms to verify authenticity and reproducibility, reducing confidence in predictive outcomes and hindering evidence-based policymaking. Therefore, an integrated approach that improves both accuracy and data transparency is needed. The combination of machine learning and blockchain technology, characterized by immutability and traceability, provides a promising solution for establishing secure and verifiable forest fire prediction frameworks.

Blockchain offers a decentralized, tamper-proof architecture that ensures data integrity through cryptographic hashing and distributed consensus mechanisms. Several studies (Datta & Sinha, 2023), (Hindarto, Hariadi, et al., 2025) have shown that blockchain enhances transparency and accountability in digital environmental data management. Nonetheless, most blockchain applications are limited to carbon emission tracking, land certification, and supply chain management, with few implementations in machine learning-based disaster prediction. The integration of these two technologies holds significant potential to create predictive systems that are both accurate and verifiable. Decentralized verification mechanisms enable each prediction to be validated independently, strengthening confidence and data reliability.

This study proposes an integrative framework combining Random Forest and XGBoost algorithms with blockchain-based data validation to predict forest fire risk in tropical regions. The case study focuses on forested areas in Cianjur Regency, West Java, Indonesia. The framework involves three stages: (1) collecting and preprocessing geospatial and environmental data using Google Earth Engine (GEE); (2) training Random Forest and XGBoost models to generate spatial fire risk predictions; and (3) validating outputs via blockchain using Ethereum smart contracts and the InterPlanetary File System (IPFS) (Zou et al., 2023). Each result is transformed into a cryptographic hash (SHA-256) and stored in a decentralized ledger, ensuring immutability, transparency, and verifiability (Zhao et al., 2022).

The novelty of this study lies in the synergistic integration of machine learning and blockchain for geospatial-based forest fire prediction. Unlike prior research focused solely on algorithmic optimization, this study introduces a dual-layer validation system in which machine learning generates spatial predictions and blockchain ensures authenticity and permanence. The research objectives are to: (1) develop a geospatial dataset relevant to fire risk; (2) build and evaluate Random Forest and XGBoost models; (3) implement blockchain as a verification mechanism. Theoretically, this work advances interdisciplinary integration between artificial intelligence and blockchain in environmental informatics, while practically, it provides a secure, transparent, and efficient model for supporting mitigation and early warning systems for tropical forest fires.

LITERATURE REVIEW

Research on forest fire prediction systems continues to evolve alongside advances in machine learning technology and blockchain-based data security. Various approaches have been taken to improve prediction accuracy and ensure the integrity of environmental data used in the analysis process. This section reviews recent research relevant to the topic of "Integrating Machine Learning and Blockchain for Forest Fire Risk Prediction" and highlights the research gaps that still need to be addressed by this study.

Machine Learning for Forest Fire Prediction

Various recent studies show that ensemble learning algorithms such as Random Forest (RF) and Extreme Gradient Boosting (XGBoost) (Hindarto, 2024a) have superior capabilities in modeling forest fire risk. (Zou et al., 2023) found that XGBoost is effective in classifying fire vulnerability levels based on geospatial variables such as Normalized Difference Vegetation Index (NDVI), ground surface temperature, rainfall, and distance to human activity. Meanwhile, (Karurung et al., 2025) shows that RF produces high accuracy in tropical regions with extreme climate variability. (Firmansyah et al., 2024) emphasizes that RF is more resistant to noise and data imbalance than conventional methods, and provides a good interpretation of the contribution of each environmental variable. However, most of these studies still focus on the accuracy of predictions and have not paid much attention to the security and authenticity of the model results.

Geospatial Data Processing and Evaluation

The quality of geospatial data plays an important role in ML model performance.. (Gupta & Kim, 2024), (Hindarto, 2024) highlights the importance of data preprocessing such as cloud masking, spatial resolution equalization, and selection of relevant vegetation indices (NDVI, NDWI, LST) before the model is trained. Research. (Yfantidou et al., 2023) shows that integrating topography, soil moisture, and distance to settlements improves the model's ability to distinguish high-risk areas. However, challenges such as class imbalance and overfitting remain common obstacles that can affect the reliability of prediction models. Most studies still use satellite data directly without a decentralized source-based validation mechanism, which can lead to the risk of data manipulation or loss.

Blockchain for Data Security and Transparency

Blockchain is known as a transparent, secure, and immutable ledger data storage technology. (Zhao et al., 2022) emphasizes that blockchain can maintain the integrity of geospatial data through cryptographic hashing and

*name of corresponding author

distributed consensus. (CHAFIQ et al., 2024), (Marzuki et al., 2025) Developing a hybrid on-chain/off-chain architecture, where big data is stored in the InterPlanetary File System (IPFS), while metadata is stored on the blockchain. This approach enables public verification without burdening the main network. (Datta & Sinha, 2023) as well as (Hindarto, Hariadi, et al., 2025) shows that blockchain can improve accountability between institutions in environmental data management through the implementation of smart contracts. However, challenges such as transaction costs (gas fees) and network scalability still limit its application in real-time systems.

In the context of implementation, (Cahyo & Hindarto, 2025) (Iswahyudi et al., 2023)explains that the Ethereum platform supports the execution of smart contracts through the Ethereum Virtual Machine (EVM). Developers can use Web3.js and the MetaMask digital wallet to connect applications to the blockchain network. Meanwhile, (Lin et al., 2022) (Hindarto, Rachmadi, et al., 2025)shows that Ganache provides a local testing environment for smart contract simulation, which helps test transaction efficiency before implementation on the public network.

Integration of Machine Learning, Blockchain, and Research Gap

The integration of machine learning with blockchain is still a relatively new area of research, especially in the context of forest fire risk prediction. Most previous studies have focused on improving model accuracy. (Zou et al., 2023), (Karurung et al., 2025), (Firmansyah et al., 2024) without considering the verification and authenticity aspects of the prediction results. On the other hand, blockchain research (Zhao et al., 2022), (CHAFIQ et al., 2024), (Hindarto, Hariadi, et al., 2025) still focuses on carbon emission tracking, land certification, and supply chain management, rather than on machine learning-based spatial prediction systems.

The research gap that has emerged can be summarized as follows: (1) Lack of verification mechanisms for blockchain-based fire prediction results. Most studies only display prediction results without cryptographic mechanisms to ensure their authenticity. (2) There is no systematic integration between ML pipelines and smart contracts; model training and prediction result storage are still carried out separately without a direct connection to the blockchain. (3) There is a lack of decentralized systems for tracking data provenance; no research has linked geospatial metadata, model parameters, and prediction results to a publicly verifiable ledger.

This research aims to bridge these gaps by proposing an integrative framework that combines RF and XGBoost for spatial prediction, as well as prediction result validation through Ethereum-based blockchain and IPFS. This approach not only improves model accuracy but also ensures transparency, integrity, and comprehensive data traceability in the tropical forest fire prediction system.

METHOD

This section presents the research methodology adopted to develop an integrated framework for forest fire risk prediction by combining machine learning, geospatial analytics, and blockchain technology. The methodology was designed to ensure systematic data processing, empirical validation, and reproducible outcomes in accordance with established quantitative research standards. A quantitative experimental approach was employed, involving the acquisition and pre-processing of multi-source geospatial datasets, the training and evaluation of predictive models, and the implementation of a blockchain-based validation mechanism. Each methodological stage—ranging from data collection, model training, and accuracy assessment to blockchain verification—is described in detail to ensure transparency and replicability. This structured methodological framework aims to produce a scientifically robust, transparent, and verifiable system for predicting forest fire risks in tropical ecosystems.

Google Earth Engine

Google Earth Engine (GEE) is a cloud-based geospatial processing platform that enables large-scale analysis of satellite imagery and environmental data. It provides access to extensive datasets, including land surface temperature, vegetation indices (NDVI), rainfall, and soil moisture, which are essential for environmental monitoring and predictive modeling. Through its powerful computational infrastructure, researchers can efficiently extract, process, and analyze multi-temporal geospatial information without requiring local hardware resources. In this study, GEE was used to collect and preprocess environmental parameters that influence forest fire risk, such as vegetation density and climatic variability. The platform's integration with machine learning workflows supports spatial prediction and visualization of high-risk areas with remarkable precision. Overall, Google Earth Engine serves as the analytical backbone of the forest fire prediction framework, ensuring efficient, accurate, and scalable geospatial data processing.

The extraction of the forest fire dataset using Google Earth Engine (GEE) was conducted through a systematic process of collecting and combining multi-source geospatial data. GEE's cloud-based environment enabled efficient access to satellite imagery and environmental variables relevant to fire risk analysis, such as land surface temperature, rainfall, humidity, and vegetation indices (NDVI). Data from platforms like MODIS, Landsat 8, and Sentinel-2 were utilized to capture temporal and spatial variations across forested regions prone to fires. Through GEE's built-in JavaScript and Python APIs, spatial filtering and region-of-interest (ROI) selection were performed to extract data specifically from the Cianjur Regency area in West Java. Each environmental parameter was processed and exported as a structured dataset in CSV format for further use in machine learning model training.

*name of corresponding author

This extraction process not only ensured high data accuracy and consistency but also supported reproducible and scalable environmental analysis for predicting forest fire risk.

Blockchain

The blockchain system functions as a distributed digital ledger that records transactions or data entries across multiple computers in a network known as nodes. Each node stores an identical copy of the ledger, ensuring that every update made to the blockchain is simultaneously reflected throughout the network. This decentralized structure eliminates the need for a central authority, reducing the risk of data manipulation and increasing transparency. In the context of forest fire prediction, the blockchain serves as a secure storage layer where environmental data, model outputs, and verification results are permanently recorded. Every new record—such as the output of Random Forest or XGBoost—is grouped into a block, cryptographically hashed, and appended to the existing chain, forming a tamper-proof sequence of data.

At the heart of blockchain's security lies cryptographic hashing, a process that transforms raw data into a unique fixed-length code. Any attempt to alter a recorded entry changes its hash value, instantly invalidating the block and alerting the network to potential tampering. This property guarantees data immutability, meaning once the data is written into the blockchain, it cannot be modified or deleted. The hash functions, such as SHA-256, play a crucial role in linking each block to the previous one, forming a continuous and verifiable data chain. This makes blockchain particularly suitable for scientific and environmental data systems, where integrity and traceability are essential.

The consensus mechanism is another foundational component that ensures all participating nodes agree on the state of the ledger. Consensus algorithms such as Proof of Work (PoW), Proof of Stake (PoS), or other variants enable the network to validate and approve new transactions collectively, maintaining consistency without requiring a central governing body. In private or research-oriented blockchain networks, lightweight consensus mechanisms are often employed to achieve faster validation with lower computational costs. Through this distributed agreement process, blockchain ensures that only legitimate and verified data are added to the ledger, reinforcing trust in the system's recorded information.

The smart contract acts as an automated and programmable protocol that executes predefined rules when certain conditions are met. In this study's context, smart contracts were developed on the Ethereum platform to automatically validate and store forest fire prediction results generated by machine learning models. Once a new prediction is produced, the smart contract records the associated data—such as coordinates, timestamp, and risk level—and generates a cryptographic signature to confirm authenticity. This automation minimizes human intervention and prevents potential bias or errors in the data entry process. Together, the ledger, nodes, consensus mechanism, and smart contracts create a self-verifying, transparent, and secure data ecosystem that strengthens accountability in forest fire monitoring and prediction systems.

The SHA-256 (Secure Hash Algorithm 256-bit) is a cryptographic hashing function used to generate a fixed-length, 64-character hexadecimal string that uniquely represents digital data. It works by converting any input—such as text, file, or dataset—into an irreversible and unique hash value. Even the slightest modification in the input data produces a completely different hash output, ensuring data integrity and immutability. In the forest fire prediction system, SHA-256 is applied to transform prediction outputs, such as risk classifications and coordinates, into unique digital fingerprints before storing them on the blockchain. This process guarantees that every record can be independently verified without exposing the original data, maintaining both transparency and security. By combining cryptographic strength and computational efficiency, SHA-256 serves as a core mechanism that upholds the reliability and trustworthiness of blockchain-based environmental data management.

Interplanetary File System (IPFS)

The InterPlanetary File System (IPFS) is a decentralized file storage and sharing protocol designed to distribute data across multiple nodes rather than relying on a single central server. Unlike traditional web systems that use location-based addressing (such as URLs), IPFS adopts content-based addressing, where each file is identified by a unique cryptographic hash derived from its content. This means that once data is stored, even the smallest change alters its hash, ensuring strong data integrity and immutability. In the context of the forest fire prediction system, IPFS is used to store large files such as environmental datasets, model outputs, and spatial risk maps generated by Random Forest and XGBoost models. Only the hash of these files is recorded on the blockchain, while the actual data remains distributed across IPFS nodes, optimizing storage efficiency and reducing blockchain load. This architecture not only guarantees that the stored data cannot be modified or deleted but also allows users to access and verify the authenticity of prediction results through a decentralized network, enhancing transparency, scalability, and reliability in environmental data management.

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

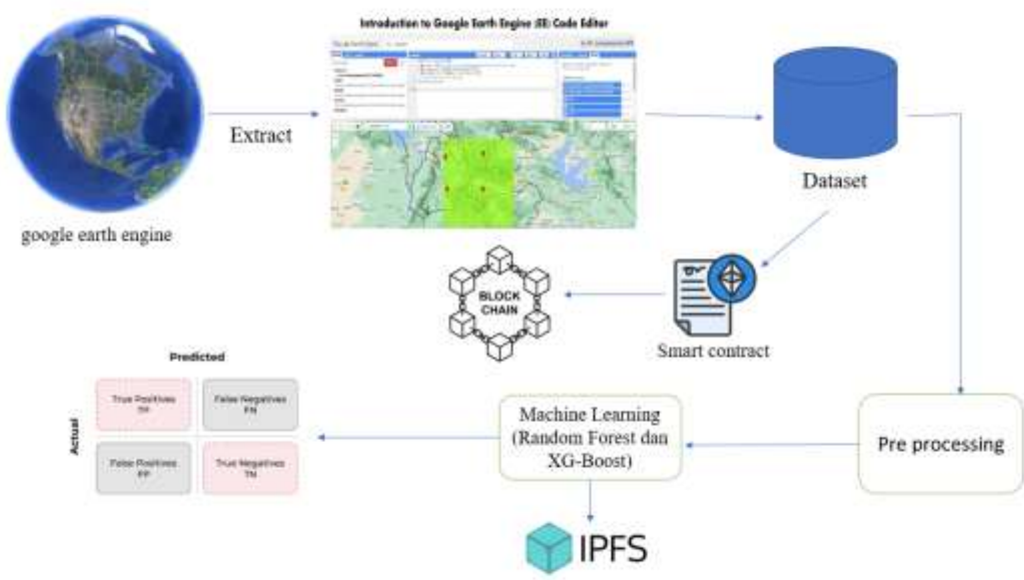

Fig 1. Research flow

Figure 1 illustrates the integrated research methodology that combines machine learning, geospatial data, and blockchain for a tropical forest fire risk prediction system. The process begins with the collection of geospatial data through the Google Earth Engine (GEE) platform, which provides satellite imagery and environmental data such as Normalized Difference Vegetation Index (NDVI), land surface temperature, rainfall, and soil moisture. The data is extracted based on specific research areas, such as the Cianjur forest area in West Java, and then stored in the form of raw geospatial datasets. NDVI is calculated using the following formula:

$$NDVI = \frac{NIR-RED}{NIR+RED} \tag{1}$$

where NIR is near-infrared reflectance and RED is red reflectance. This NDVI value reflects the level of vegetation density and is an important variable in determining the level of fire risk. The extracted dataset is then stored and used as the basis for spatial analysis and machine learning model training.

The next step is data preprocessing, which aims to clean, normalize, and standardize the scale of all variables so that they can be processed by machine learning algorithms. This process includes removing outliers, filling in missing values, and normalizing numerical data using the Min–Max Normalization method, with the formula:

$$X' = \frac{X-X_{min}}{X_{max}-X_{min}} \tag{2}$$

This formula is used to convert the value of each feature into a range of 0–1 so that each variable has an equal weight. After preprocessing, the dataset is divided into training data (70%) and testing data (30%), then trained using two main algorithms, namely Random Forest and XGBoost, to produce a fire risk prediction model.

Random Forest

The Random Forest model proved effective in identifying forest fire risk levels, particularly for regions with low fire probability. It performed consistently in recognizing dominant environmental patterns such as temperature, humidity, and wind speed, which are crucial for understanding the dynamics of fire-prone areas. However, the model tends to prioritize majority data classes, resulting in lower sensitivity to minority classes like medium- or high-risk zones. This bias is typical of ensemble models that aggregate many decision trees, where frequent patterns tend to dominate predictions. Still, its ability to handle large datasets and complex nonlinear relationships makes it a dependable choice for modeling environmental data.

In practice, Random Forest offers an appealing balance between performance, interpretability, and efficiency. It is relatively easy to implement, requires minimal parameter tuning, and provides valuable insights into feature importance—helping researchers understand which variables most strongly influence fire occurrence. Its robustness to noise and outliers allows it to deliver stable predictions even when data quality varies across regions. Consequently, Random Forest serves as a strong foundational model for environmental risk assessment systems, offering a reliable benchmark before advancing to more complex algorithms like XGBoost.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15373
e-ISSN : 2541-2019
p-ISSN : 2541-044X

In the Random Forest algorithm, decision trees are formed randomly by selecting subsets of data and features for each decision tree. The final prediction process is based on a majority vote of all trees, which can be written mathematically as:

$$\hat{y} = \text{mode}\{h_1(\text{x}), h_2(\text{x}), \ldots, h_k(\text{x})\} \quad (3)$$

XGBoost

The XGBoost model demonstrated superior adaptability and precision in predicting forest fire risk. It effectively captured complex nonlinear interactions between environmental variables, allowing it to identify high-risk areas that other models might overlook. By learning iteratively from its own prediction errors, XGBoost continuously refined its accuracy, resulting in more balanced predictions across all classes. The model's ability to handle class imbalance and its sensitivity to subtle variations in climate data made it particularly strong in distinguishing between medium- and high-risk zones. This strength reflects XGBoost' s gradient boosting mechanism, which focuses learning power on difficult cases to improve overall generalization.

Beyond its predictive strength, XGBoost also excels in efficiency and scalability, making it suitable for large geospatial and climatological datasets. Its interpretability through feature importance visualization enables researchers to understand which environmental parameters—such as temperature, humidity, and rainfall—most strongly influence fire risk. This insight is invaluable for developing targeted fire prevention strategies and early warning systems. Moreover, when integrated into the Forest Fire Dapps framework, XGBoost enhances the reliability of blockchain-recorded predictions, ensuring transparent and traceable results. In essence, XGBoost not only improves predictive accuracy but also strengthens the system's scientific credibility and trustworthiness.

In XGBoost, the training process is carried out iteratively using a gradient boosting approach to minimize the loss function with model regulation, which is formulated as:

$$\text{Obj}(\text{t}) = \text{i} = \sum_{i=1}^{n} l\left(\text{y}_i, \hat{\text{y}}_i^{(\text{t}-1)} + \text{f}_t(\text{x}_i)\right) + \Omega(\text{f}_t) \quad (4)$$

with l as the loss function, ft as the new decision tree function at iteration t, and Ω as the regularization parameter to prevent overfitting.

The result of this model is a confusion matrix consisting of True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN). From this, performance metrics such as accuracy, precision, and F1-score are calculated, which are formulated as:

$$\text{Accuracy} = \frac{\text{TP+TN}}{TP+TN+FP+FN} \quad (5)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (6)$$

$$\text{F1} - \text{Score} = 2 \times \frac{\text{Precision}\times\text{Recall}}{\text{Precision+Recall}} \quad (7)$$

The model with the highest F1-score is considered the best model and is used to generate a fire risk classification map with three categories: Low Risk, Medium Risk, and High Risk.

The final stage in the diagram shows the integration of prediction results with blockchain technology, where the model output (risk prediction, execution time, and location coordinates) is recorded in hash form using the SHA-256 (Secure Hash Algorithm 256-bit) cryptographic algorithm. This hash function is written as:

$$\text{H} = \text{SHA} - 256(\text{x}) \quad (8)$$

where H is the unique hash value of data x (risk map prediction results). The hash is stored in a smart contract on the Ethereum Blockchain network, while the prediction results file is stored decentralized on the InterPlanetary File System (IPFS). The smart contract functions as an automatic verification mechanism to ensure the integrity and authenticity of the prediction results data, so that each prediction result can be re-verified without being manipulated. With this approach, the system produces an accurate, efficient, and secure workflow, and supports the principles of data transparency and accountability in tropical forest fire mitigation.

RESULT

This section of the research presents empirical findings from the geospatial data analysis process using Random Forest and XGBoost algorithms that have been integrated with a blockchain-based validation system. The main focus of this section is to showcase the model's performance in predicting the spatial and temporal risk levels of tropical forest fires. In addition, the validation test results and accuracy comparisons between algorithms are explained systematically to demonstrate the effectiveness of the integrative approach used. Overall, this section provides a quantitative and visual basis that supports the interpretation and discussion in the next section.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15373
e-ISSN : 2541-2019
p-ISSN : 2541-044X

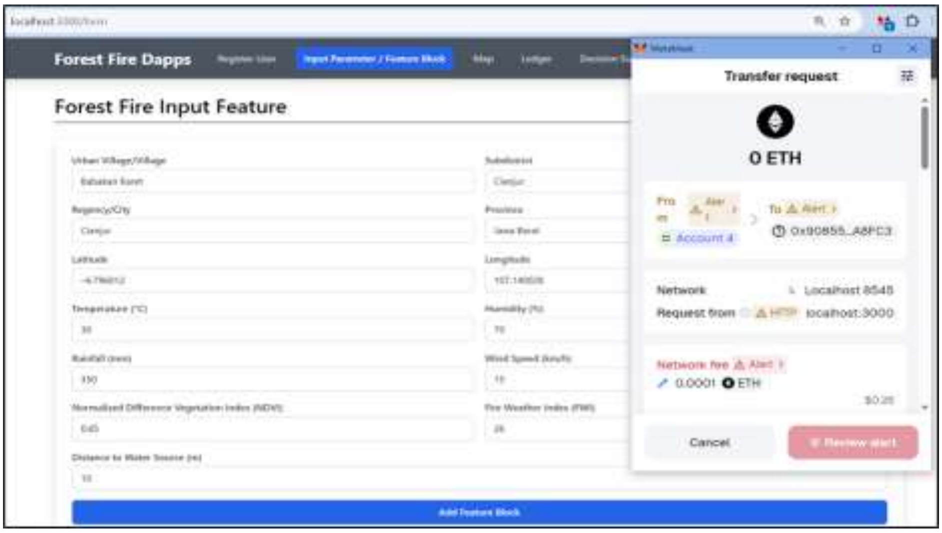

Fig 2 forest fire Dapps system display

Figure 2 shows the interface of the Forest Fire Dapps; a decentralized app built to manage and record forest fire risk parameters via blockchain technology. On the left, users can enter environmental and spatial data like temperature, humidity, rainfall, wind speed, NDVI, and location coordinates for areas such as Babakan Karet, Cianjur. After clicking the "Add Feature Block" button, the system connects to MetaMask, securely storing the data as cryptographically locked blocks on the Ethereum blockchain. Each transaction incurs a small network fee and generates a unique digital signature to prevent data tampering or deletion. The data is then analyzed by Random Forest or XGBoost models to predict fire risk levels. This integration ensures transparent, verifiable, and tamper-proof predictions, combining artificial intelligence and blockchain technologies. reliable forest fire mitigation.

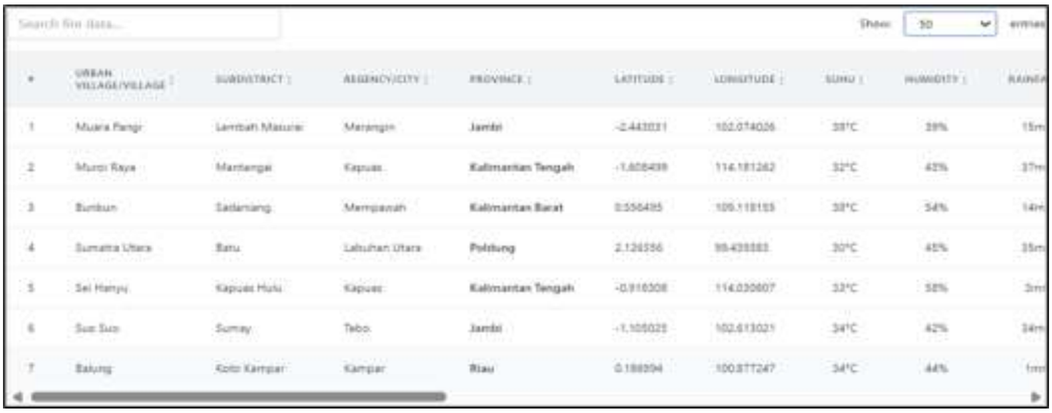

Fig 3 Ledger parameter fire forest

Figure 3 illustrates the data table interface of the Forest Fire Dapps system, which displays environmental parameter inputs from various forest fire–prone areas across Indonesia. Each row represents a village or location—such as Muara Pangi, Muroi Raya, and Bunbun—along with administrative data like subdistrict, district, and province. The table includes key variables such as latitude, longitude, temperature, humidity, and rainfall, which serve as crucial indicators for assessing fire vulnerability. For instance, Muara Pangi's high temperature and low humidity reflect a high fire risk, while Sei Hanyu's higher humidity indicates a lower one. This integrated dataset supports the Random Forest and XGBoost models in predicting fire risk based on complex climatic and geographical interactions. Moreover, all entries are recorded on the blockchain through smart contracts to guarantee data transparency, authenticity, and immutability.

The Random Forest model achieved an overall accuracy of 88.3%, showing strong performance in classifying low-risk wildfire regions with near-perfect recall and precision. However, it struggled slightly with the minority classes representing medium and high risk. The model's macro precision of 0.88 and macro recall of 0.72 indicate that while it performs well on average, it tends to favor the majority class. This is common in imbalanced datasets, where the model becomes slightly conservative in predicting rarer events like high-risk fire zones. Still, Random

*name of corresponding author

Forest proved effective as a baseline classifier, capturing the general trends of the environmental variables influencing wildfire risk.

The XGBoost model surpassed the Random Forest with an accuracy of 90%, demonstrating better balance between precision (0.86) and recall (0.78) across all classes. Its macro F1-score of 0.81 reflects more stable learning across both frequent and rare classes, suggesting that XGBoost captured complex feature interactions more effectively. While both models performed well, XGBoost' s gradient boosting mechanism allowed it to refine predictions iteratively, improving the detection of medium- and high-risk zones. This makes XGBoost the more robust choice for predicting wildfire risk patterns in diverse environmental conditions.

Table 1. Result training Random Forest and XGBoost

Model	Accuracy	Precision (Macro)	Recall (Macro)	F1-Score (Macro)
Random Forest	0.883	0.88	0.72	0.76
XGBoost	0.900	0.86	0.78	0.81

Table 1, interpretation both models demonstrate strong predictive performance, but XGBoost slightly surpasses Random Forest in balancing precision and recall across all classes. Random Forest performs exceptionally well for the majority class, effectively identifying low-risk areas, yet it struggles to accurately classify minority classes with higher fire risk. XGBoost, on the other hand, manages class imbalance more effectively, resulting in higher macro recall and F1-score values. This indicates that XGBoost captures complex patterns in the data more comprehensively, enhancing its generalization capability. In practical applications, XGBoost is better suited for detecting rare or high-risk fire cases due to its robustness and adaptability. However, Random Forest remains a valuable option for rapid development and preliminary analysis because of its simplicity and faster training time.

DISCUSSIONS

The results of this research show that both Random Forest (RF) and XGBoost algorithms are capable of predicting forest fire risk effectively based on geospatial and climatological data. The Random Forest model successfully captured general environmental patterns such as temperature, humidity, and rainfall, allowing it to identify low-risk regions with strong consistency. However, it showed limited sensitivity to medium and high-risk categories due to class imbalance, a common issue in ensemble tree-based methods. Despite this, Random Forest remains valuable as a stable and interpretable baseline model for understanding variable contributions in fire risk prediction. Its outputs demonstrate that environmental conditions with high temperature and low humidity strongly correlate with increased fire susceptibility.

In contrast, the XGBoost model performed with better balance and precision across all risk categories. Its gradient boosting mechanism allowed for iterative refinement of predictions, effectively managing nonlinear interactions and class imbalance. The model produced more accurate classifications of medium- and high-risk areas by learning from previous errors, improving generalization performance. The results reveal that XGBoost is highly responsive to subtle variations in climate variables and vegetation indices, enabling it to identify critical high-risk zones more effectively. This confirms that advanced ensemble techniques provide not only higher prediction accuracy but also better adaptability in diverse environmental conditions.

The integration of blockchain technology added a new layer of transparency and reliability to the forest fire prediction framework. Each model output—whether from Random Forest or XGBoost—was transformed into a cryptographic hash and recorded on the Ethereum blockchain through smart contracts. This ensured that every prediction was immutable and verifiable, preventing any unauthorized modification of data. The combination of artificial intelligence and blockchain thus forms a secure and decentralized "verifiable prediction pipeline." This system supports data accountability and enhances confidence in predictive analytics for forest fire mitigation, providing a solid foundation for transparent and data-driven environmental management.

CONCLUSION

Forest fires remain a major environmental challenge in tropical regions such as Indonesia, where vast forest ecosystems are increasingly vulnerable to climate change and human activities. The rising frequency of droughts, land clearing, and inadequate monitoring systems have intensified fire occurrences, causing severe ecological and socioeconomic impacts. Addressing this issue requires accurate and transparent fire risk prediction systems that can support early mitigation and decision-making. This study introduced an integrative framework combining machine learning and blockchain technology to enhance both predictive accuracy and data reliability. Using geospatial and climatological data extracted from Google Earth Engine (GEE), two ensemble algorithms—Random Forest (RF) and XGBoost—were trained to model forest fire susceptibility. The results showed that Random Forest effectively recognized dominant environmental patterns but tended to favor majority classes, while XGBoost produced more balanced and adaptive predictions across all risk levels. The inclusion of blockchain

*name of corresponding author

added a decentralized verification layer through smart contracts and cryptographic hashing, ensuring the integrity and traceability of each model output. These findings demonstrate that the integration of artificial intelligence and blockchain creates a verifiable prediction pipeline for secure and transparent fire risk monitoring. This approach addresses the limitations of traditional centralized systems prone to data manipulation and loss. Overall, the study concludes that this framework provides a robust, efficient, and trustworthy solution for improving the accuracy, transparency, and accountability of tropical forest fire risk prediction systems.

REFERENCES

Badan Nasional Penanggulangan Bencana. (2024). Data dan Informasi Kebencanaan Bulanan Teraktual. *Info Bencana*, *5*(1), 1–19. https://bnpb.go.id/storage/app/media/Buletin Info Bencana/Buletin Info Bencana Januari 2024.pdf

Cahyo, F. Y. N., & Hindarto, D. (2025). Smart Contract Architecture for a Blockchain-Driven Multi Criteria DSS in Forest Fire Monitoring and Response. *Sinkron*, *9*(3), 1146–1158. https://doi.org/10.33395/sinkron.v9i3.15009

CHAFIQ, T., AZMI, R., Fadil, A., & Mohammed, O. (2024). Investigating the potential of blockchain technology for geospatial data sharing: Opportunities, challenges, and solutions. *Geomatica*, *76*(2), 100026. https://doi.org/10.1016/j.geomat.2024.100026

Datta, S., & Sinha, D. (2023). BSEIFFS: Blockchain-secured edge-intelligent forest fire surveillance. *Future Generation Computer Systems*, *147*, 59–76. https://doi.org/10.1016/j.future.2023.04.015

Firmansyah, A., Syahidin, M. F., & Triana, Y. S. (2024). Prediksi Kebakaran Hutan Berdasarkan Titik Panas dan Iklim Menggunakan Algoritma Random Forest. *Jurnal Nasional Teknologi Dan Sistem Informasi*, *10*(2), 145–155. https://doi.org/10.25077/teknosi.v10i2.2024.145-155

Gupta, R., & Kim, H. (2024). A Geospatial Approach to Wildfire Risk Modeling Using Machine Learning and Remote Sensing Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *17*, 13570–13576. https://doi.org/10.1109/JSTARS.2024.3434368

Hindarto, D. (2024a). Case Study: Gradient Boosting Machine vs Light GBM in Potential Landslide Detection. *Journal of Computer Networks, Architecture and High Performance Computing*, *6*(1), 169–178. https://doi.org/10.47709/cnahpc.v6i1.3374

Hindarto, D. (2024b). Edge Computing Architecture Sensor-based Flood Monitoring System: Design and Implementation. *Sinkron*, *8*(3), 1758–1769. https://doi.org/10.33395/sinkron.v8i3.13874

Hindarto, D., Hariadi, M., & Rachmadi, R. F. (2025). A Blockchain-based Landslide Mitigation Recommendation System for Decision-Making. *Engineering, Technology and Applied Science Research*, *15*(1), 20452–20460. https://doi.org/10.48084/etasr.9806

Hindarto, D., Rachmadi, R. F., Hariadi, M., & Damastuti, F. A. (2025). Contextual Awareness System for Landslide Risk Recommendation in Crypto-Spatial. *2025 International Electronics Symposium (IES)*, 700–706. https://doi.org/10.1109/IES67184.2025.11161195

Iswahyudi, I., Hindarto, D., & Indrajit, R. E. (2023). Digital Transformation in University: Enterprise Architecture and Blockchain Technology. *Sinkron*, *8*(4), 2501–2512. https://doi.org/10.33395/sinkron.v8i4.12977

Karurung, W. S., Lee, K., & Lee, W. (2025). Assessment of forest fire vulnerability prediction in Indonesia: Seasonal variability analysis using machine learning techniques. *International Journal of Applied Earth Observation and Geoinformation*, *138*(February). https://doi.org/10.1016/j.jag.2025.104435

Lin, S., Li, Z., Zhao, S., Zhao, H., Li, Y., & Wang, S. (2022). Design and Implementation of Blockchain-based College Education Integrity System. *2022 IEEE 5th International Conference on Information Systems and Computer Aided Education (ICISCAE)*, 276–281. https://doi.org/10.1109/ICISCAE55891.2022.9927601

Marzuki, I., Hindarto, D., Dzikri, A., Damastuti, F. A., Arif, Y. M., Rachmadi, R. F., & Hariadi, M. (2025). Hierarchical Clustering-Based Geospatial Analysis for a Personalized Tourism Destination Recommender System. *Engineering, Technology and Applied Science Research*, *15*(4), 24794–24799. https://doi.org/10.48084/etasr.11055

Yfantidou, A., Zoka, M., Stathopoulos, N., Kokkalidou, M., Girtsou, S., Tsoutsos, M. C., Hadjimitsis, D., & Kontoes, C. (2023). Geoinformatics and Machine Learning for Comprehensive Fire Risk Assessment and Management in Peri-Urban Environments: A Building-Block-Level Approach. *Applied Sciences (Switzerland)*, *13*(18). https://doi.org/10.3390/app131810261

Zhao, P., Cedeno Jimenez, J. R., Brovelli, M. A., & Mansourian, A. (2022). Towards geospatial blockchain: A review of research on blockchain technology applied to geospatial data. *AGILE: GIScience Series*, *3*, 1–6. https://doi.org/10.5194/agile-giss-3-71-2022

Zou, Q., Yu, W., & Bao, Z. (2023). A Blockchain Solution for Remote Sensing Data Management Model. *Applied Sciences (Switzerland)*, *13*(17). https://doi.org/10.3390/app13179609

*name of corresponding author