[1] **Naga Gopi Chappidi**
[2]**Nizampatnam Yashwanth**
[3] **Karri Snehalatha Reddy**
[4] **Gudapati Snohitha Sai Sri**
[5] **Dr. Dubba Naga Malleswari***

Blockchain and Machine Learning Synergy: An Approach to Decentralized and Secure Model Training

Journal of Electrical Systems

Abstract: - Through the past few years machine learning and blockchain technologies have rapidly advanced and revolutionized numerous industries. Blockchain offers secure transaction tracking, whereas machine learning facilitates data-driven decision making. This study explores both technologies and their potential to transform finance, supply chain management, healthcare and identity management. The focus is on how this integration can improve cybersecurity, privacy, and implement decentralization nature through improved data management and analysis. This collaboration promises automation, transparency, and data-driven decision-making while strengthening security. However, challenges, such as security threats, integration planning, and data processing, must be addressed before widespread adoption. Surmounting these hurdles will unlock the full capabilities of this technological fusion, shaping the future of enhanced efficiency, security and data-driven decision making across various sectors.

Keywords: Machine learning, Decentralization, Blockchain, Smart Contract, Immutable ledger, Identity management, Finance, Transparency, Secure model training, Digital voting system.

I. INTRODUCTION

ML redefines computers learning. Instead of following rigid instructions, ML algorithms delve into the data and uncover hidden patterns. This empowers them to make predictions or decisions on new data, without explicit programming. ML has invisibly transformed our daily lives from spam filtering to facial recognition. **"**The concept of blockchain was first proposed by Satoshi Nakamoto in 2008 through the use of a consensus protocol**" [1]**. **"**The blockchain operates as a secure and decentralized digital record-keeper, documenting transaction details such as date, time, price and participating parties**" [2]**. Blockchain aims to build trust and safety through its traits of keeping data intact, ensuring security, being dependable, and spreading control across many users. These characteristics collaborate to enhance the system's reliability. The synergy of BT and ML generates a transformative advantage ushering in a new era of data management and analysis. Blockchain's secure and immutable ledger evolves into a dependable data repository, enabling the effective use of extensive datasets by ML algorithms. This offers a significant advantage for tasks involving sensitive data, as the cryptographic foundation of the blockchain guarantees data integrity and traceability. This transformative capability will not only enhance data analysis but also facilitates predictive modeling, driving innovation across diverse sectors. This enhanced analysis strengthens security by identifying fraudulent transactions or vulnerabilities within the BN.

Even more groundbreaking, this synergy can lead to the onset of autonomous smart contracts. This paves the way for smart contracts to learn and adapt their execution based on real-world data feeds, potentially revolutionizing data management, security, and decision-making across a multitude of industries.

1.1 Abbreviations

ML – Machine learning
BT – Blockchain technology
BN – Blockchain network

[1,] Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur – 522302, Andhra Pradesh, India. 2100039020cse.r@gmail.com
[2] Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur – 522302, Andhra Pradesh, India. 2 2100039119cse.r@gmail.com
[3] Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur – 522302, Andhra Pradesh, India. 3 2100030245cser@gmail.com
[4] Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur – 522302, Andhra Pradesh, India. 4 2100030185cser@gmail.com
[5] Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur – 522302, Andhra Pradesh, India.
* Corresponding Author Email: nagamalleswary@kluniversity.in

1.2 Literature Review

Initially, we searched standard databases, including IEEEXplore, ScientificReports, ScienceDirect, Springerlink, and Google Scholar, to locate existing research studies. We employed keywords such as ML and BT, Integration of AI and blockchain, blockchain and machine learning applications during our search. We identified numerous studies, but we classified and selected a few to work on. Ultimately, we chose three exceptional papers to serve as the foundation of our research as depicted in Table 1.

The remaining sections of the paper are structured as following; In Section 2, we introduce BT and ML. In Section 3, we address the limitations of ML and strategies for overcoming these limitations using the BT. Section 4, highlights the applications and challenges of this integration. In Section 5, we demonstrated a case study on decentralized and secure model training with Blockchain, and the conclusion, future enhancements of our research had presented in Section 6.

Table 1: List of papers

Refs.	Paper title	Year	Objective
[3]	Blockchain meets machine learning: a survey	2024	the transformative impact of blockchain and machine learning across diverse sectors, paving the way for groundbreaking advancements and innovative possibilities.
[4]	Survey on the Convergence of Machine Learning and Blockchain	2022	address issues related to data privacy, model updating, traditional machine learning and strategies.
[5]	Machine Learning Adoption in Blockchain-Based Smart Applications: The Challenges, and a Way Forward	2020	challenges and solutions of integrating machine learning into blockchain-based smart applications. Discover how data analytics can enhance security and privacy in blockchain technology.

II. BLOCKCHAIN TECHNOLOGY

"BT was originally introduced by Haber and Stornetta in the early 1990s as a means of securing timestamps and connecting digital data in a linear chronological sequence.**" [6]**. Blockchain operates as a decentralized ledger comprising a chain of interconnected records or blocks that grow over time. These blocks are securely linked through cryptographic hashes. Each block includes crucial findings such as a cryptographic hash of the preceding block, timestamp, and transaction data, often organized in a Merkle tree format. This structure creates a chain-like sequence, akin to a linked list, where each new block is connected to a previous one. Consequently, altering data in a single block would necessitate changing all subsequent blocks, rendering blockchain transactions irreversible. **"**The elimination of the need for a central authority in the database structure is one of the most significant and powerful features of the blockchain.**" [7]**. In today's networking landscape, decentralized systems have gained prominence, wherein all computers within a network collaborate, without the necessity for a central control point. This approach puts forward numerous advantages, such as efficient sharing of files and resources among computers. Unlike centralized setups, decentralized networks provide enhanced reliability by mitigating the risk of a single failure source. These are meant to be the main focus of the public ledger in the form of distributed technology. The network comprises a distributed network of nodes that collectively follow a consensus protocol to incorporate and confirm new transactions. The network utilizes a distributed computing framework called Byzantine fault tolerance. This approach ensures dependability and robustness against malicious intrusions and system breakdown.

2.1 Peer to peer network

In a peer-to-peer (P2P) system, individual computers or devices known as nodes interact directly with one another without the need for a centralized server as depicted in Figure 1. This indicates that all nodes have equal power and can act as both clients and servers.

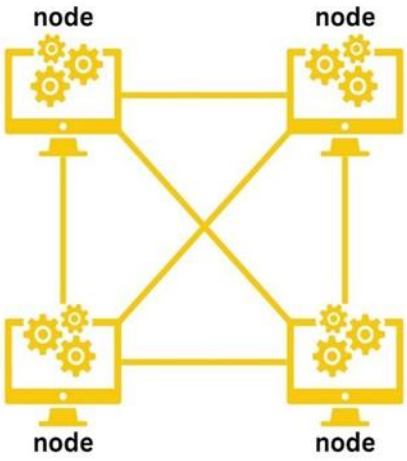

Figure1. Representation of p2p network

Each node in a P2P network has a complete copy of the blockchain. This guarantees that all participants share the same transaction history, removing the necessity for a central authority to validate it. Because there is no central server that acts as a lone point of failure, tampering with data becomes extremely challenging. Any modifications to the data must be reflected in all copies of the data across network making them highly unlikely to go unnoticed. Even though the size and type of information that blocks can store may vary based on the system's design, a cryptographic hash function is employed to link the blocks and create a complex mathematical relationship between them. Figure 2 shows the block representation in the BN.

1. **Data:** The recorded information.
2. **Hash:** a cryptographic function that converts random data into a fixed length output.
3. **Previous hash:** Blocks are connected through the previous hash of the preceding block, which safeguards against the modification or insertion of blocks between existing ones.
4. **4.Timestamp:** relates to recording of the time a particular transaction or event occurred in the BN.

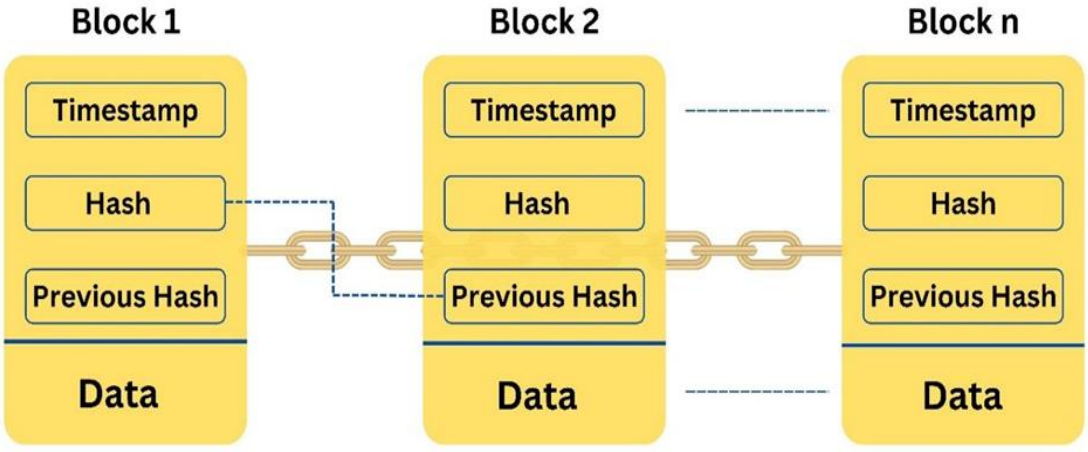

Figure 2. Block representation

2.1.1 Merkle tree

A Merkle tree is a fundamental data structure essential for effectively summarizing and validating the integrity of extensive datasets. It is generated through a process of repeatedly combining and hashing pairs of data elements until a single root hash is obtained as depicted in Figure 3 **[8]**. In this structure every leaf node represents a data element, while the intermediate nodes contain the cryptographic hash of their respective child nodes. In blockchain systems, the Merkle tree is indispensable for verifying the validity of the transactions and blocks. The Merkle root enables network nodes to efficiently verify the inclusion of specific transactions within a block without processing the entire block's contents. This method enhances BN's scalability and security while reducing the computational power required for validation, all without compromising data's immutability.

2.1.2 Smart contracts

Smart contracts function as digital intermediaries in the BT. These automated programs encode the conditions of an agreement, and upon fulfillment of all specified conditions, they facilitate the exchange of assets (such as money, goods, and information) **[9]**. For example, when making an online purchase, a smart contract can safeguard payment until the item is delivered, and then it can be released to the seller automatically. By eliminating reliance on a third-party intermediary, smart contracts provide a secure and transparent solution, with all details of the agreement publicly visible on the blockchain. Smart contracts hold immense promise for streamlining processes, reducing friction in transactions, and fostering trust assurance in a digital world. As smart contract functionalities continue to evolve, we can prospect even broader applications to emerge across various industries. Combining smart contracts with ML can accelerate the smart contract process by optimizing and preprocessing the required actions. ML algorithms predict and resolve bottlenecks, ensuring a smoother and faster contract execution. This integration allows dynamic adjustments and streamlined operations, resulting in more responsive and efficient systems. Consequently, complex transactions are swiftly and effectively handled.

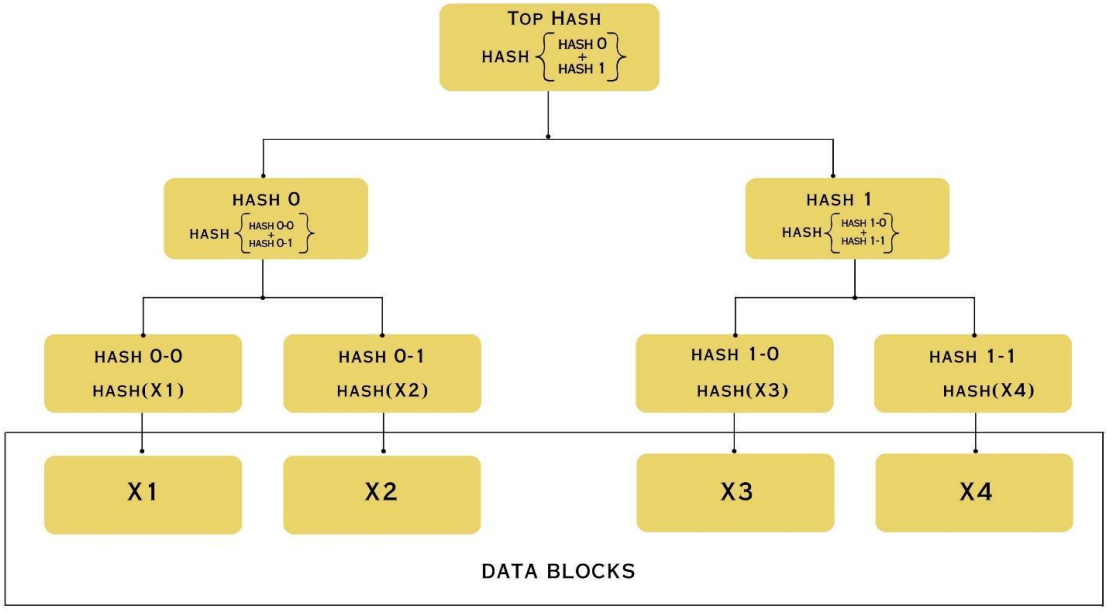

Figure 3. Merkle tree

2.2 Machine learning

"Machine learning, a subfield of artificial intelligence, focuses on creating algorithms and methodologies that allow computers to learn from data and make predictions or decisions without human intervention**" [10]**. By identifying patterns within the data with which they interact, these systems gradually enhance their capabilities and performance through continuous refinement over time. **"**The idea of machine learning was initially introduced by Arthur Samuel, an American computer scientist, in 1959. It refers to the ability of a computer to acquire knowledge without being explicitly programmed**" [11]**. ML excels at making predictions in complex scenarios. Unlike conventional methods that depend on hand-coded rules, ML systems learn from past data. By examining data and identifying patterns, they acquire the capability to make predictions on new and unseen data. The accuracy of these predictions relay on the quality and quantity of data used to train the model. This approach leverages powerful algorithms which is able to trained using large datasets. These algorithms can then autonomously discover the logic for making predictions based on the inherent patterns within the data. Figure 4 shows simplified representation of the operation of a ML algorithm. ML encompasses various techniques for training algorithms, as depicted in Figure 5 **[12]**

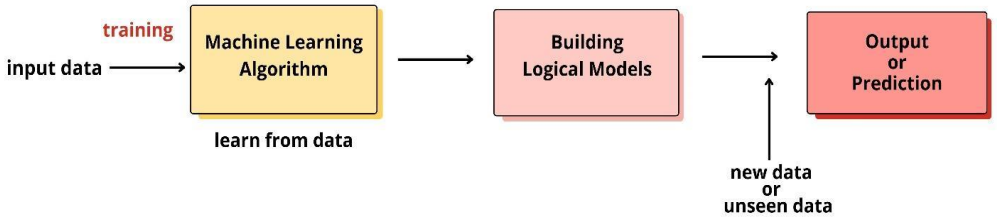

Figure 4. Operation of ML algorithm

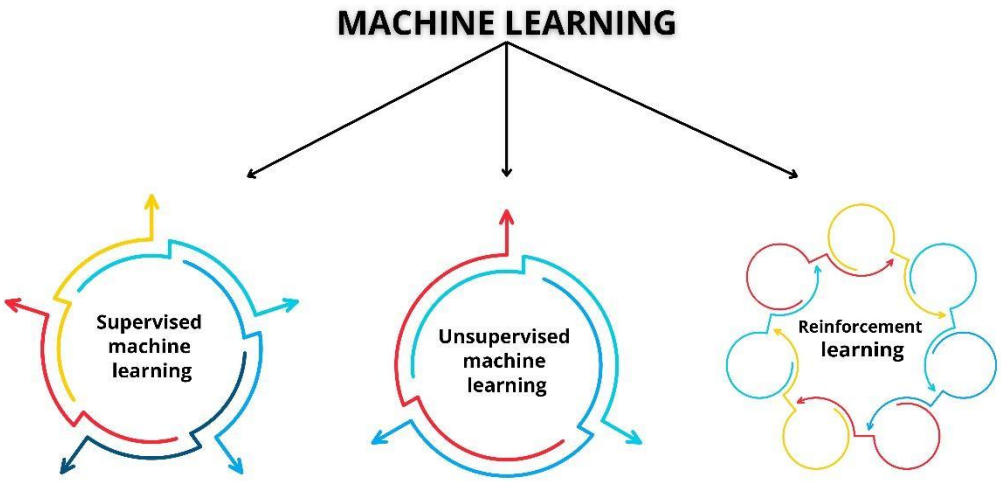

Figure 5. Categories of ML

2.2.1 Supervised learning

Supervised learning carves a path into two distinct approaches: classification and regression. These methods allow machines to gain encounters by accessing labeled data, with each data point acting as a guide with both an input and a corresponding desired output.

Classification: Classification tasks involve assigning new data points to pre-established categories. Imagine a system meticulously sorting incoming emails into "important" and "spam" folders. Classification algorithms were

employed to distinguish the abstract ideas and connections not further the training data, enabling efficient categorization of unseen data into appropriate classes. Examples of classification applications include spam filtering, image recognition for object identification in photos, and fraud detection in financial transactions.

Regression: Regression concentrates on identifying numerical rates, like evaluating major costs based on circumstances such as location and size along with predicting weather patterns. Mainly these algorithms gain information from data, each point of data involves an integer value aimed at the variable that the algorithm learns to anticipate. Regression methods find broad application across different domains, including analyzing stock markets, forecasting sales to predict future figures, and predicting real estate prices.

Supervised learning models (SLM), involve linear regression, Support Vector, Logistic Regression, Decision Trees which are utilized in tasks where labeled training data enables algorithms to predict outputs based on input data.

2.2.2 Unsupervised learning

Unsupervised learning applies unlabeled data in which information lacks predefined categories. Imagine a basket of seashells, all unique yet unlabeled. **"**Unlike supervised learning, unsupervised techniques do not require pre-sorting, instead they analyze the data to uncover hidden patterns and groupings**" [13]**. This allows them to identify similarities, like clustering seashells based on color or texture. Unsupervised learning excels in data exploration and discovery, providing valuable insights even without specific prediction goals.

Clustering: Clustering algorithms are adept at categorizing data points according to their inherent similarities. Imagine sorting seashells in the basket we mentioned earlier. Clustering can group them according to color, texture, or even size. This facilitates the identification of natural groupings within the data, uncovering latent structures and relationships that may not be readily evident.

Dimensionality Reduction: Dimensionality reduction aims to capture the most important information from complex data, even if it has a lot of variables. Imagine that each seashell has information about its weight, thickness, and other details. Dimensionality reduction simplifies data by keeping only the most relevant features. It condenses a vast amount of information into a smaller set of key characteristics, making it easier to analyze and visualize the data.

Unsupervised Learning models consist of K-means Clustering, Hierarchical Clustering, Principal Component Analysis (PCA), and Apriori Algorithm for Association Rules. These models are employed in tasks where the data is unlabeled, and the algorithms learn to find the hidden patterns or intrinsic structures in the input data.

2.2.3 Reinforcement learning

Reinforcement learning (RL) carves a distinct path from its supervised and unsupervised counterparts. While supervised learning thrives on meticulously labeled data and unsupervised learning excels at uncovering hidden patterns within unlabeled data, RL leverages an interactive learning paradigm. This method resembles how living organisms learn by exploring and interacting with their surroundings, where actions are assessed on the basis of rewards or penalties they encounter.

2.2.4 Core Concepts in Reinforcement Learning:

Agent environment Interaction: RL centers around an agent, an entity capable of taking actions within a dynamic environment. The environment responds to these actions by providing the agent with rewards (positive reinforcement) or penalties (negative reinforcement) on the basis of outcome. These rewards or penalties guide the learning process of the agent.

Feedback-Driven Learning: Through a process of ongoing trial and error, the agent discovers the connection between its actions and the resulting rewards. This iterative feedback loop is vital, as it helps the agent improve its behavior over time and aim to optimize its overall reward in the given environment.

Policy Optimization: The core objective of an RL agent is to foster an optimal policy and set of rules that dictate its actions within the environment. Through reinforcement learning algorithms, the agent iteratively updates and improves its policy to maximize the long-term reward accumulation.

Reinforcement Learning models encompass Q-Learning, Deep Q Networks (DQN), Policy Gradient Methods, and Actor-Critic Models. These models are employed in situations where an agent gains decision-making skills by interacting with an environment, with the goal of maximizing its cumulative reward over time.

2.2.5 Deep learning

As a branch of ML, Deep learning replicates the neural architecture and operations of the human brain. It employs multilayered artificial neural networks with numerous connections to evaluate and process data. Each layer transforms the data, building towards a final output namely image recognition or prediction. Unlike simpler models, deep learning displays hundreds of layers, allowing the extraction of intricate patterns from vast datasets. This learning using an example approach fuels its continuous improvement. Deep learning's impact stretches across various fields, from powering facial recognition in smartphones to enabling natural language processing in chatbots. DL remarkable capacity to process intricate data and continuously learn, deep learning has swiftly transformed numerous industries and is significantly influencing the future of artificial intelligence.

III. LIMITATIONS AND STRATEGIES

3.1 Limitations of machine learning

- ML models can be vulnerable to manipulation through the introduction of inaccurate or misleading information into training data. This "data poisoning" can skew the model's learning process and lead to compromised security outcomes.
- Malicious actors can exploit vulnerabilities in ML models by crafting specifically designed inputs (adversarial examples) that cause the model to make incorrect predictions. This can be particularly dangerous in security applications, where compromised models can overlook threats.
- Unauthorized acquisition or access to a trained ML model by unauthorized parties poses a significant security risk. For instance, a compromised model used in a security system can be manipulated to bypass security measures.
- Unintentional biases present within the training data could be inadvertently incorporated into the ML model. This algorithmic bias may lead to security vulnerabilities if the model executes the discriminatory decisions that compromise the security protocols.

3.2 Strategies

BT shows potential in addressing numerous challenges inherent in ML, particularly those related to security and transparency.

First, the blockchain's immutable ledger ensures the integrity of stored data and protects against tampering or poisoning. The use of BT and smart contracts assures that the training dataset is stored securely and accessible only to authorized persons. This approach controls both contributing and retrieving data from the dataset.

Second, blockchain can authenticate data inputs and outputs, and detect discrepancies caused by adversarial attacks through timestamping and secure storage of model predictions. Consensus mechanisms further aid in identifying and rejecting malicious input.

Third, storing trained models on a blockchain protects its integrity and prevents unauthorized access or tampering. Smart contracts can govern access rights, restrict deployment, and interact with authorized parties.

Fourth, blockchain-based decentralized data marketplaces facilitate the acquisition of diverse and unbiased training data, with recorded metadata enabling transparency in data sourcing and model training processes. This allows stakeholders to audit and verify the fairness of a model.

Finally, the transparent nature of the blockchain enables stakeholders to trace data and model updates, thereby enhancing accountability and trust. In addition, cryptographic techniques provide verifiable proof of the model behavior, thereby increasing interpretability and explainability.

By integrating BT with MT, organizations can bolster the security, transparency, and fairness of AI applications, thereby mitigating inherent risks.

IV. Applications And Challenges

2.3 Applications

2.3.1 Supply Chain Management

The current supply chains are experiencing a transformation propelled by the integration of ML and BT. This synergy ensures new levels of transparency and traceability throughout the process. Businesses are empowered to safeguard product authenticity and quality by preventing counterfeiting and fraud effectively. ML algorithms act as intelligent analysts and filter through massive datasets within the supply chain. By uncovering patterns, anomalies, and areas appropriate for improvement, these algorithms pave the way for optimized operations and data-driven decision making. The BT serves as a secure foundation for this transformation. It functions as an incorruptible ledger, conscientiously recording every transaction and movement of goods. This allows businesses to meticulously track products in detail from their origin to their final destination, thus creating immutability and transparency. **“**The DHL Global Trade Barometer is an example of a software application that integrates AI and blockchain technologies to enhance supply chain management **“[14].**

2.3.2 Finance

“This integration enhances fraud detection, risk assessment, and trading strategies in the financial sector**” [15]**. Blockchain's unchangeable record boosts security and transparency, while ML algorithms scrutinize vast datasets to identify patterns and anomalies. This collaboration enables financial institutions to enhance detection of fraudulent activities and make well-informed, data-driven choices. For instance, blockchain encryption safeguards transactional privacy and data integrity, while ML models continuously adjust to market changes. An example of this integration is the application of homomorphic encryption methods to maintain privacy during model training. By encrypting data before processing with ML algorithms, sensitive information remains protected throughout analysis.

2.3.3 Health care

By combining blockchain's immutable ledger with ML algorithms, healthcare providers can securely store and analyze vast volumes of patient data **[16]**. This integration not only safeguards data integrity but also facilitates the extraction of valuable insights for personalized treatments and predictive analytics. For instance, BT can facilitate the creation of decentralized patient health records, empowering individuals to manage their medical data securely while upholding their privacy. Meanwhile, ML algorithms can parse through these data and identify patterns, anomalies, and potential disease trends, thereby enhancing diagnosis accuracy and treatment effectiveness. Collaborative efforts across the healthcare sector are driving advancements in this domain, with partnerships among technology firms, healthcare providers, and research entities fueling innovation. Despite their considerable

potential benefits, challenges such as data standardization and regulatory compliance require attention to foster widespread adoption.

2.3.4 Identity management

The integration of ML and BT into identity management enhances security, privacy, and user control over personal data. ML algorithms analyze biometric and behavioral data to securely verify digital identities, whereas blockchain ensures the integrity, transparency, and decentralization of identity records. For example, consider digital voting system, a government agency adopts a digital voting system that combines ML and BT to enhance voter authentication and election integrity. ML algorithms analyze biometric data, voter behavior patterns, and historical voting records to securely verify voter identities. BT encrypts and stores anonymized voting records, ensuring the transparency and auditability of election results, while protecting voter privacy.

2.3.5 Internet of things (IoT)

The incorporation of ML and BT into IoT security improves cybersecurity, data integrity, and privacy protection for the connected devices and networks **[17]**. By utilizing ML algorithms to scrutinize IoT data for anomalies and security threats, and leveraging BT for secure communication, authentication, and data sharing, the integration of these technologies significantly bolsters IoT system security. For instance, in a smart city setup, the municipal administration adopted a smart city infrastructure that incorporated both ML and BT to enhance cybersecurity and data privacy. ML algorithms evaluate sensor data from IoT devices throughout the city to identify anomalies, forecast traffic congestion, and optimize energy usage. Meanwhile, BT secures communication channels between IoT devices and data storage systems, preventing unauthorized access and ensuring data integrity.

2.4 Challenges

Computational Bottlenecks: Balancing the processing demands of complex ML algorithms with the ledger maintenance requirements of BT can create limitations. Optimizations through specialized hardware like GPUs and exploring alternative blockchain consensus mechanisms with lower computational needs are potential solutions.

Data Security and Privacy: Secure storage of sensitive data on a blockchain is achievable through hashing. However, additional measures, such as encryption, might be necessary, as hashing itself can sometimes reveal information about the data format.

Scalability and Sustainability: Current blockchain limitations include scalability and energy consumption as the user base and transactions increase. Future research must focus on developing scalable blockchain architectures and exploring greener consensus mechanisms to confront these concerns.

Interoperability and Communication: Standardized data formats and communication protocols are crucial to bridge the gap between different BT and ML systems, enabling seamless integration.

Smart Contract Security: To minimize the potential impacts of flaws or mistakes in smart contracts, it is crucial to implement thorough code reviews, employ formal verification methods, and adopt secure programming practices.

Data Quality and Trust: ML algorithms rely heavily on data quality. While blockchain's secure and tamper-proof data storage enhances trust in data integrity, ensuring the quality and accuracy of data fed into ML models remains paramount.

Collaborative Decision-Making: Blockchain serves as a reliable and transparent platform for data storage and collaboration, facilitating collaborative decision-making processes in which stakeholders can access and verify information on the blockchain.

Transaction Costs: Transaction fees associated with certain blockchain platforms can be considered. Traditional databases may be more cost-effective for small-scale projects that have limited data and updates.

Data Ownership and Accountability: Blockchain's possibility to redefine data ownership is substantial, as it offers users with enhanced control over their personal information. Furthermore, blockchain's data provenance and lineage tracking capabilities can address concerns surrounding accountability in the "black-box" nature of ML.

V. CASE STUDY

To demonstrate this integration, we aimed to explain it using a case study as illustrated in Figure 6.

Step 1: Data collection and Pre-processing module

The module begins by gathering data from a variety of sources. These data serve as the raw material for the modeling process, and once collected, the data undergo pre-processing. This crucial step cleans and prepares data for modeling tasks. Preprocessing techniques address issues such as inconsistencies, missing values, and formatting errors, thereby ensuring that the data are suitable for analysis.

Step 2: Feature engineering and Model training module

Feature engineering techniques are applied to pre-processed data and then the chosen ML model was trained on prepared data in a decentralized training environment. This environment should prioritize security measures mainly access control and intrusion detection.

Step 3: Model hashing and Metadata generation module

Once trained, the model is hashed using a robust cryptographic hash function (e.g., SHA-256) to compose a unique fingerprint. This module creates comprehensive metadata associated with the model training process as follows:

a) Generating model hash value
b) Training data provenance details (source, timestamps, cryptographic hashes of the facts used for training)
c) Model training parameters and configuration details (hyperparameters, algorithm used)
d) Cryptographic signatures from authorized participants involved in training

Step 4: Blockchain integration module

This module interfaces with the chosen blockchain platform to secure the generated metadata. The specific implementation depends on the selected blockchain and its functionality. The following are two potential approaches.

a) On-chain Storage: If the blockchain platform allows for sufficient storage capacity and doesn't qualify as major concern, the entire model hash and critical metadata can be stored directly within the blockchain.
b) Off-chain Storage with On-chain Anchoring: For larger models or cost limitations, the model itself can be stored off-chain in a reliable and tamper-proof storage system (e.g., hardware security modules). The model hash and the reference (pointer) to its secure storage location can be anchored to the BN for verification.

Step 5: Model testing and Deployment module

Before the model's full-scale deployment, it was extensively tested using a separate dataset that was not employed during the training process. This testing was conducted to discover possible issues such as overfitting or poor generalizability. Upon successful testing, the trained model was deployed on a server for real-world predictions. To safeguard against unauthorized access or tampering with the model, it is essential to establish strong security protocols.

Step 6: Prediction and Verification module

Users interact with the deployed model to predict the data. The deployed model can integrate logic to verify its integrity, as follows:

a) Recalculating its hash and comparing it with the hash stored in the blockchain.
b) Potential fetching of additional verification information from the blockchain, such as cryptographic signatures from authorized participants involved in training.

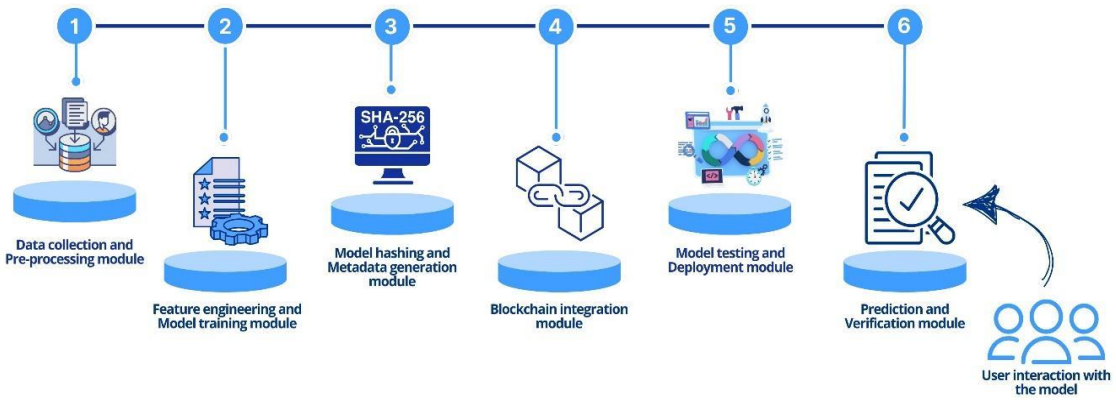

Figure 6. Decentralized and Secure Model Training with Blockchain

This case study provides a framework for leveraging blockchain to secure a ML model training process. It offers enhanced data provenance, tamper resistance, and potential for improved accountability. However, careful consideration of scalability, computational overhead, and blockchain platform selection is crucial for a successful implementation.

VI. CONCLUSION

In this study, we aimed to explore how the synergy between BT and ML improves the ML model training. We also explored the key areas where integration was used, and found that the synergy between BT and ML presents a multitude of career opportunities across various domains. This survey has illuminated the potential of leveraging the blockchain's decentralized architecture and immutability with the predictive power and data analysis capabilities of ML. By enhancing transparency, security, and efficiency, this integration can streamline complex processes and foster trust between participants, paving the way for a more decentralized and user-centric Web 3.0. However, challenges such as scalability, interoperability, and privacy concerns necessitate collaborative efforts by researchers, developers, and policymakers. Evolving regulatory frameworks that encourage innovation while mitigating risk is crucial. Despite these hurdles, the potential benefits are profound. Through continued interdisciplinary collaboration, strategic investments in research and development(R&D), and open knowledge sharing, we can unlock the full potential of this powerful combination by harnessing the synergy between BT and ML. We can forge a new path that will shape the future of Web 3.0 and driving positive changes across various sectors. This future web promises greater user control over data, a more level-playing field for creators and consumers, and a more secure and transparent online experience.

6.1 Future research and enhancements

Although the synergy between BT and ML offers exciting possibilities, future research must address these key challenges. We need to improve scalability to meet the demands of ever-growing ML tasks. Compatibility between different blockchains and ML frameworks is crucial for wider adoption. In addition, developing robust privacy-preserving techniques, such as homomorphic encryption, is essential for securing data on blockchains. Regulatory frameworks must evolve to accommodate decentralized technologies, and ethical considerations should guide research to ensure fairness and accountability of AI models. Finally, the focused exploration of industry-specific applications in healthcare, finance, IoT and supply chains through collaboration and pilot projects is indispensable for driving real-world innovation that prioritizes cost-effectiveness. In the future, the convergence of BT and ML will create opportunities for specialized roles such as Decentralized AI Architects and Quantum Blockchain Engineers. Ethical AI Validators will ensure fairness in algorithms, while Blockchain Integration Specialists will facilitate seamless adoption across industries. Interdisciplinary collaboration, such as Smart Contract Auditors and Distributed Ledger Analysts, will play key roles in assuring the integrity and efficiency of blockchain-based systems. Experts skilled in navigating the changing terrain of this field will propel innovation and mold the future of work in this dynamic profession.

REFERENCES

[1] Nakamoto S. Bitcoin whitepaper. https://bitcoin.org/bitcoin.pdf; 2008.

[2] Zheng, Zibin, Shaoan Xie, Hongning Dai, Xiangping Chen and Huaimin Wang. “Blockchain challenges and opportunities: a survey.” Int. J. Web Grid Serv. 14 (2018): 352-375.

[3] Kayikci, S., Khoshgoftaar, T.M. Blockchain meets machine learning: a survey. J Big Data 11, 9 (2024). doi: 10.1186/s40537-023-00852-y.

[4] Ding, Shengwen & Hu, Chenhui. (2022). Survey on the Convergence of Machine Learning and Blockchain.

[5] S. Tanwar, Q. Bhatia, P. Patel, A. Kumari, P. K. Singh and W. -C. Hong, "Machine Learning Adoption in Blockchain-Based Smart Applications: The Challenges, and a Way Forward," in IEEE Access, vol. 8, pp. 474-488, 2020, doi: 10.1109/ACCESS.2019.2961372.

[6] Haber, S., Stornetta, W.S. How to time-stamp a digital document. J. Cryptology 3, 99–111 (1991). doi: 10.1007/BF00196791

[7] M., Niranjanamurthy & Nithya, B. & St, Jagannatha. (2019). Analysis of Blockchain technology: pros, cons and SWOT. Cluster Computing. 22. 10.1007/s10586-018-2387-5.

[8] Wikipedia. https://en.wikipedia.org/wiki/Merkle_tree.

[9] B. K. Mohanta, S. S. Panda and D. Jena, "An Overview of Smart Contract and Use Cases in Blockchain Technology," 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Bengaluru, India, 2018, pp. 1-4, doi: 10.1109/ICCCNT.2018.8494045.

[10] Mitchell, T. M. (1997), Machine learning, Vol. 1, McGraw-hill New York.

[11] Samuel AL. Machine learning. Technol Rev. 1959;62(1):42–5.

[12] Javatpoint. https://www.javatpoint.com/machine-learning.

[13] Vats, Varun & Zhang, Lining & Chatterjee, Sreejit & Ahmed, Sabbir & Enziama, Elvin & Tepe, Kemal. (2018). A Comparative Analysis of Unsupervised Machine Techniques for Liver Disease Prediction. 486-489. 10.1109/ISSPIT.2018.8642735.

[14] DHL Global Trade Barometer. https://lot.dhl.com/global-trade-barometer-gtb/.

[15] T. H. Pranto, K. T. A. M. Hasib, T. Rahman, A. B. Haque, A. K. M. N. Islam and R. M. Rahman, "Blockchain and Machine Learning for Fraud Detection: A Privacy-Preserving and Adaptive Incentive Based Approach," in IEEE Access, vol. 10, pp. 87115-87134, 2022, doi: 10.1109/ACCESS.2022.3198956.

[16] Goyal, Adit & Elhence, Anubhav & Chamola, Vinay & Sikdar, Biplab. (2021). A Blockchain and Machine Learning based Framework for Efficient Health Insurance Management. 511-515. 10.1145/3485730.3493685.

[17] Chowdary, Ch & Puli, Srilakshmi & K, Lakshmi & Santhi, M.V.B.T. (2021). Machine Learning Based Data Security Model Using Blockchain for Secure Data Transmission in IoT. 1521-1527. 10.1109/ICESC51422.2021