International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Integrating Blockchain and AI for Data Encryption and Secure ETL Pipelines

Manohar Reddy Sokkula

Submitted: 12/01/2025 **Revised:** 26/02/2025 **Accepted:** 17/03/2025

***Abstract*—**In the era of data-driven decision-making, ensuring the security, transparency, and integrity of Extract, Transform, and Load (ETL) pipelines has become increasingly critical, especially in regulated industries such as healthcare, finance, and telecommunications. Traditional ETL systems often rely on centralized architectures with basic encryption and access control mechanisms, which, although essential, fall short of addressing sophisticated cyber threats, data tampering, and compliance verification. This research proposes a hybrid framework that integrates Blockchain technology and an MLP-GRU (Multi-Layer Perceptron – Gated Recurrent Unit) neural network to enhance the security and intelligence of ETL processes. Blockchain is employed to create a decentralized, tamper-proof ledger that logs each ETL operation, providing traceability, immutability, and auditability. In parallel, the MLP-GRU model is utilized to detect anomalies in ETL activities by analyzing both static and sequential log data. This dual approach ensures not only secure data management but also real-time monitoring and predictive threat mitigation. The experimental setup involves blockchain-based logging of ETL operations and AI-based anomaly detection, evaluated using metrics such as Accuracy (99%), Precision (98.21%), Recall (98%), and F1-Score (98.77%). Results demonstrate that the integrated system outperforms traditional ETL security mechanisms in detecting malicious activity while maintaining efficient data throughput and low latency. Furthermore, the study examines blockchain transaction performance under varying data volumes to validate the scalability of the proposed solution. The framework's ability to automate compliance verification and generate immutable audit trails presents a significant advancement in secure data pipeline design. Future work includes enhancing privacy through Zero-Knowledge Proofs, scaling to federated systems, and incorporating advanced deep-learning architectures. Overall, this research sets a strong foundation for the development of intelligent, secure, and regulation-compliant ETL infrastructures through the convergence of blockchain and AI technologies.

***Keywords*—.** *Blockchain, ETL Pipeline, MLP-GRU, Anomaly Detection, Data Security, Compliane, Artificial Intelligence.*

I. INTRODUCTION

Today, in a hyperconnected digital economy, tremendous quantities of heterogeneous data are generated, processed, and transferred for an organization to operate and compete [1]. Enterprises are relying on advanced ETL pipelines to combine various data sources into actionable intelligence for anything ranging from real-time financial transactions and claims from an insurance company to the health records of a patient and usage logs of a telecom operator. In other words, these pipelines enable processes such as integration, quality improvement, semantic transformations, and optimized storage for analytics and decision-making. But with data growing in volume and complexity and its very nature becoming sensitive, their conventional ETL infrastructures' vulnerabilities have become glaringly evident. Classic ETL systems are typically built on centralized architectures. As such, they present highly attractive opportunities for cyberattacks and insider threats, raising considerable risks concerning unauthorized access, tampering of data, loss of lineage, and lack of verifiability.

Although traditional ETL pipelines have been subject to encryption using SSL/TLS, firewalls, role-based access controls, and other conventional security measures, they could never be truly deemed traceable and hence immutable, ensuring data integrity throughout the entire data lifecycle [2]. Conventional security measures imply a basic level of defense but are mostly reactive and hence insufficient against modern cyber threats, from a so-called "advanced" persistent attack, all the way to insider threats. The main challenge emanates from the centralized architecture that most ETL systems subscribe to data is passed through single processing nodes and finally stored in central warehouses, something highly susceptible to malicious tampering, corruption, and sudden catastrophic failure if the system gets compromised. These centralized systems do not provide it with verifiable audit trails, those trails that can independently verify whether data has been altered, copied, or erased illegally, thereby putting a heavy blackout on transparency and accountability.

By offering a decentralized and tamper-proof ledger recording transactions and operations in a transparent and

Sr. Solutions Architect, Corpay

verifiable manner, blockchain technology is set to transform solutions to the entrenched issues with data security and data integrity posed by the traditional ETL pipelines [3]. At the core of blockchain are cryptographic algorithms and consensus mechanisms working in tandem to time-stamp, uniquely hash, and irrevocably bind each data interaction in a chain of prior history- everything is there to prevent any unauthorized compromise. If appropriately embedded in ETL processes, the blockchain backbone can be used to keep an auditable record of every significant step in the ETL life cycle: extraction, transformation, and loading. This record of activities is kept forever, enabling one to guarantee the data's authenticity and data traceability; further, this late record could give them real-time verification and the possibility for useful forensic analysis when any data tampering, unauthorized access, manipulation, or data forgery attempts are being notified [4].

While blockchain technology provides a firm technical guarantee of security, immutability, and auditability of data workflows, AI stands to greatly augment the functionality and flexibility of ETL pipelines by endowing them with cognitive abilities, scalability, and real-time decision-making power. AI-fed systems continually scrutinize vast volumes of ledger entries and metadata generated during ETL operations to rapidly detect abnormalities, forecast potential security threats, and intelligently assign computational resources to ensure optimal performance [5]. Deploying AI models along with machine learning and deep learning algorithms can allow independent adaptive behavior toward changing data structures, schema variation management, and complex data cleaning with maximum accuracy while reducing manual interventions and processing delays. Subsequently, an AI analytical engine catalogs subtle irregularities or anomalies in data pipelines, warranting tamper alerts, data drift, or system failure, thus either triggering a remediation or escalation protocol [6].

The fusion of blockchain and AI into ETL processes gives birth to a hybrid framework that augments data security while maximizing operational intelligence and flexibility. Blockchain technologies, including smart contracts, cryptographic hashing, and distributed consensus, find their way into every stage of the ETL pipeline in this synergistic architecture, thereby ensuring that every data transaction, transformation, and access event is recorded permanently, with tamper evidence and with cryptographic verification. Such blockchain-based technologies thereby provide an incorruptible, decentralized, transparent audit trail to ensure data integrity, uptime for compliance, and removal of silent data distortion at will [7]. Concomitantly, AI engines survey the ETL stand, analyze patterns, recognize anomalies, and intervene in system threats or performance degradation in real time, utilizing advanced techniques like machine learning, natural language processing, and neural networks. These intelligent agents adapt themselves dynamically to changes in data sources, optimize transformation rules, and raise alarms against dubious activities or attempts of unauthorized access [8].

The generation of a hybrid structure by fusion makes it relevant for domains that are dealing with data sensitive to the highest degree, critical to life momentarily, or data intensively regulated. In the healthcare sector, since patient records are extremely private and must be accurate, the blockchain can timestamp and trace all interactions with the data from diagnosis through treatment updates to insurance claims to ensure complete data lineage and non-repudiation, while the AI systems examine the access logs and movement of patient data to spot anomalies, unauthorized access, or compliance breaches in real-time [9]. In the financial industry, AI can detect fraud or suspicious behavior in high-frequency market transaction streams faster and more accurately than has ever been possible before, while blockchain immutably records every single transaction and audit event that takes place, thus ensuring compliance with regulations like AML and KYC [10].

Another big advantage of blockchain integration with AI is that the process has become smoother to facilitate regulatory compliance. With global attention growing on data privacy, protection, and governance, legislation such as GDPR, HIPAA, and PCI DSS, among others, applies rigorous criteria to data collection, processing, storing, and sharing. Organizations must keep verifiable and real-time audit trails to show compliance proactively. The power of blockchain is that every data transaction, transformation, and access event within an ETL process gets recorded on a decentralized and immutable ledger that is public and tamper-proof [4]. Such traceability assures that every interaction with sensitive data is recorded chronologically, cryptographically secured, and verifiable to an independent party crucial element while proving compliance in audits or legal scenarios [11]. AI, on the other hand, infuses an element of automation and smartness in managing compliance by monitoring ETL processes consistently, alerting on any breaches of policy, verifying data access controls, and comparing real-world practices to regulatory standards.

With the above-explained advantages, the research study focuses on conceptualizing, designing, and implementing the next-generation ETL frameworks with blockchain and AI synergistically to create a secure, intelligent, and resilient data processing environment [12]. At its heart, the design of the pipeline must ensure full-stage encryption and security at extraction, transformation, and load while having the cognitive capability to learn from historical data behavior, identify anomalies, and preemptively address any potential risks and vulnerabilities [13] [14]. This approach secures the data using blockchain with decentralized and immutable records that provide tamper-proof logging and data lineage verifiability and leverages AI models for situational awareness and complexity in decision-making, including deep learning, anomaly detection algorithms, and predictive analytics. The research evaluates various blockchain platforms for their suitability in secure ETL integration (e.g., Hyperledger Fabric, Ethereum, and Corda) while simultaneously researching AI-based techniques suitable for

embedding into systems for real-time monitoring and fraud detection.

The Key contributions of the article are given below,

- A novel integration of blockchain technology into the ETL pipeline ensures data immutability, tamper-proof logging, and traceability across extract, transform, and load operations, enhancing transparency and regulatory compliance.
- The proposed model intelligently detects anomalies in ETL operations by combining an MLP for feature abstraction and a GRU for temporal behavior learning, enabling real-time detection of irregular patterns or security threats.
- Smart contracts are deployed to enforce access control, workflow validation, and automated responses to anomalies, providing decentralized self-governance and improving the reliability of data operations.
- Instead of storing raw data, the system logs cryptographic hashes, timestamps, and anomaly flags on the blockchain, ensuring lightweight, cost-effective, and audit-friendly metadata storage while preserving confidentiality.

This document is organized as follows for the remaining portion: Section II discusses the related work. The problem statement is discussed in Section III. The recommended method is described in Part IV. In Section V, the experiment's results are presented and contrasted. Section VI discusses the paper's conclusion and suggestions for further study.

II. Related Works

A. ETL Systems

To handle data heterogeneity and event interpretation in intricate systems like computer networks and telephones, Cichonski et al. [15] outline an end-to-end data processing architecture that blends Semantic Web technologies with traditional NMSs and SIEMs. Semantic Web tools for knowledge representation, including provenance tracking, declarative data mapping using RML, batch and stream processing, data patching, and reconciliation based on SPARQL and SKOS, and semantic data transfer based on Kafka, are integrated into the suggested architecture, setting it apart from traditional systems. The offered architecture demonstrates its remarkable ability to combine disparate data sets for monitoring and security analytics by producing an RDF knowledge graph that can detect cross-domain irregularities in industrial environments.

The need for strong ETL processes in scenarios where digital data is becoming increasingly varied in terms of both structured and unstructured data is covered in length by Kumaran [16]. These, along with big data frameworks like Hadoop and Spark, will be increasingly useful since managing unstructured data—such as text, photos, and video content—requires more flexible AI-driven ways. Relational databases with preset schemas are usually used to process structured data using SQL-based tools. Additionally, it offers comprehensive coverage of hybrid ETL pipelines, which complement one another to deliver optimal performance and scalability analytics. It discusses several strategies to improve efficiency and integration across diverse data sources and provides best practices for resolving mixed-data ETL process problems.

B. ETL Use Cases

The typical ETL processes are about to be modified by Seenivasan [17] for usage with cloud data engineering. It fixes several problems, including resource waste, excessive latency, and mismatched data transformation. AI-driven features like intelligent workload management, automatic schema generation, and real-time anomaly detection make ETL pipelines more scalable, flexible, and efficient. It also describes how to use these advantages of AI in real-world applications that demonstrate notable improvements in data processing accuracy, speed, and overall operational efficiency. Finally, it points out that AI ETL systems are already playing a significant part in modern, high-performance data-engineering solutions in more complex and dynamic cloud infrastructures.

The Enhanced Temporal-BiLSTM Network, or ETLNet, is a model proposed by Ansari et al. [18] to detect road anomalies such as potholes and speed bumps. Instead of using visual input, which has been demonstrated to be unsuccessful in low light or unmarked regions, this model makes use of data from smartphone inertial sensors. ETLNet claims that a BiLSTM layer is combined with two TCN layers. These layers are designed to evaluate gyroscope and accelerometer data separately to identify irregularities on various road surfaces. This is a great study for creating advanced automated traffic monitoring systems that can be used in autonomous vehicles and public transportation.

C. Security Using ML

Joshi [19] examines the drawbacks of traditional batch-oriented ETL processes for managing fast, real-time data and proposes state-of-the-art machine-learning techniques to build ETL pipelines that are flexible and self-improving. Real-time ETL is enhanced by the use of predictive modeling, anomaly detection, reinforcement learning-based resource allocation, and schema drift management. Such intelligent pipelines would be able to take proactive steps to manage workloads, preserve data quality, and even adjust to changes in data architecture on their own by utilizing time series prediction and learning-based insights. Experimental validations on systems like Databricks and AWS Glue demonstrate significant benefits, including a 40% reduction in latency and a 25% reduction in resource expenditures. This study illustrates the potential for ML-enhanced ETL systems to become effective and independent.

The major security concerns that emerge in cloud and distributed systems—which can be large, flexible, and cost-effective—are the focus of Saswata Dey, Writuraj Sarma, and Sundar Tiwari [20]. These systems are also susceptible to a variety of advanced threats, including DDoS attacks, insider threats, and zero-day attacks. This elucidates how DL models, including CNNs, RNNs, and transformers,

enhanced the ability to define patterns and were able to recognize these threats instantly. Scalable cloud deployment is another consideration when handling unbalanced data and combining DL with edge computing performance improvements. The findings of the experiment show that DL models perform better than traditional methods in terms of anomaly detection and virus prevention.

D. Secure Data Transfer

Concerns have been expressed over the rising cybersecurity vulnerabilities as a result of organizations' increasing reliance on internet services, digital storage, and software-oriented processes. Proactive vulnerability assessments must be carried out since digital transformation leaves IT infrastructures vulnerable to potential threats. Thus, the objective of Hiremath et al. [21]is to identify system vulnerabilities and collect relevant information for developing effective solutions using data analytics tools like Power BI. Helping clients create a safe online environment that protects their private information from hackers is the aim.

For effective data transfer from Oracle BI into Salesforce, minimizing system interruptions, and ensuring data integrity during the transition from conventional to cloud-based systems, Hamza et al. [22] recommend an ETL-based strategy. It explains how the Extract, Transform, and Load processes may enhance operational performance and promote data mobility, especially when taking finance and ERP into account. The research presents data virtualization as a method that may be a very flexible and scalable alternative for accessing data in real-time without significant duplication to support Agile processes and expedite decision-making. The same is applied to enhance business intelligence skills and predictive analytics.

III. Research Methodology

A. Research Gap

With the rising concern about the architecture of secure data pipelines, ETL systems have been focused mainly on traditional security considerations such as encryption, access control, and rudimentary audit logging. These techniques, although useful up to a point, fail to offer a concrete set of solutions against advanced cybersecurity efforts, insider threats, and data lineage tampering [23]. Most ETL frameworks are therefore designed around a centralized architecture, inherently prone to single points of failure and limited traceability, and vulnerable to illicit changes, among others. While some attempts were made to secure the data flowing inside them, the little literature available has hardly explored the entire range of end-to-end security for ETL. There is an obvious absence of frameworks that provide decentralization along with real-time monitoring and immutable audibility, all integrated coherent and scalable manner.

The application of advanced technologies such as blockchain and AI toward ETL security remains underexplored and fragmented in the literature. Blockchain is typically discussed in the context of financial transactions or supply chain management, with the question of how the same technology should be applied to ETL pipeline integrity and audit seldom explored [24]. Likewise, AI has found varied applications in anomaly detection and optimization methods, yet rarely is its capability to improve ETL pipelines, especially when paired with blockchain, studied. Experimental evaluation of performance trade-offs and interoperability challenges while merging these two technologies within real-world, data-heavy scenarios is also missing. Hence, this research aims to fill this gap by proposing a unified intelligent and secure ETL framework that harnesses blockchain and AI toward strengthening robust data processing, end-to-end visibility, and predictive threat mitigation.

B. Proposed Framework

The layered diagram in Fig. 1 offers a representation of a blockchain and AI-empowered secure and intelligent ETL framework for anomaly detection. At the highest point in the stack lies Raw Data, which acts as the primary input and consists of unstructured or structured dataset types from different sources, including databases, IoT devices, enterprise applications, etc. The raw data then proceeds to Data Collection, during which appropriate fields are extracted and structured for processing. Data Preprocessing transforms the input, cleaning and normalizing it (Min-Max scaling, for instance) and selecting the salient features needed for the analysis to ensure consistency and readiness for modeling. Blockchain Integration subsequently records every step of the transformation, metadata information, and corresponding hashes into an immutable secure ledger, providing traceability, lineage, and tamper-proof audit trails. Anomaly detection comes next, whereby the system uses sophisticated AI models based on the MLP-GRU architecture to track temporal patterns and detect abnormal behaviors that provide indications of system faults, security intrusion, or data inconsistencies.

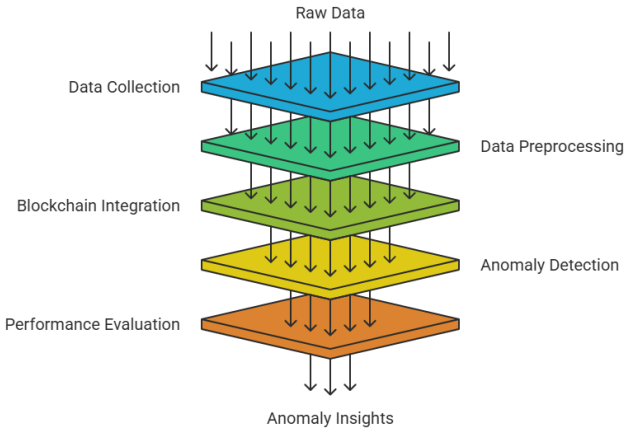

Fig. 1. Proposed Framework

C. Extract – Data Collection

Having to evaluate the proposed blockchain-integrated ETL pipeline with MLP-GRU-based anomaly detection, a

comprehensive and diversified dataset was needed; one that truly represents ETL operations in a real-world environment. Both synthetic and publicly available operational log datasets were employed for this purpose. Synthetic data configurations emulated different ETL activities regarding data extraction, transformation, and loading, both under normal and abnormal scenarios. Anomalous events were described as abrupt surges in data volume, unauthorized access attempts, transformation failures, irregular job durations, and changes in access roles. The parameters included in this dataset ranged from job ID, timestamp, user role, type of task (extract, transform, load), processing time, and data size, to system response codes. This dataset's diversity ensured the training and testing of the model with an enormous variety of patterns, which will sharpen its generalization and subtle deviation-detecting capability.

The model was also evaluated in real settings and further validated with real-world datasets generated from platforms such as Kaggle and open-source ETL tools. Logs extracted from these systems contained records arranged in a structured manner to give complete accounts of data flow activities and system behaviors. They were then preprocessed to discard irrelevant features, impute missing values, normalize numerical fields, and encode categorical variables. For the temporal modeling aspect adopted by the MLP-GRU, sliding windows were set up within time series to sufficiently represent the sequential nature of ETL operations. Labeled sequences encompassed both normal and anomalous instances so that the model could learn the patterns displaying safe versus attacked states. This mixture of synthetic data plus real data has made the training dataset balanced and contextually rich enough to fairly evaluate performance and assist in building a strong, secure ETL framework.

D. Transform - Data Preprocessing Using Min-Max Normalization

During the ETL pipeline transformation process, data preprocessing plays a great role in preparing the raw data for efficient and accurate anomaly detection. One of the important preprocessing techniques applied in the present study is Min-Max normalization, which scales all numerical features into a uniform range, typically between 0 and 1. This prevents features with larger magnitudes from dominating the learning process, thereby giving a chance to the MLP-GRU model to learn from each given input variable meaningfully. If the dataset contains measurements from heterogeneous variables such as job execution time, data size, block commit duration, and throughput values that differ enormously in scale, normalization becomes even more critical. Min-max normalization ensures consistency within the data so that all features become comparable, which in turn helps the model learn faster and converge better. It is given in Eq. (1).

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

Where:

X = Original value

X_{min} = Minimum value in the feature column

X_{max} = Maximum value in the feature column

This normalization of data preserves the integrity of patterns within temporal sequences, something critical for the GRU component to comprehend trends as they evolve. It is also advantageous for the MLP to receive inputs on a uniform range so that the training is faster and mitigation measures against computational catastrophes, such as exploding gradients, are in place. Before normalization, the dataset went through the removal of missing values and one-hot encoding of categorical variables to ensure that the entire dataset was numerical and machine-readable. Finally, the normalized data was divided into training, validation, and testing sets, with the normalization set applied to each of the three. This transformational step makes sure to standardize the dataset while making sure that the AI model can work with inputs optimized for anomaly detection, resulting in higher model performance and more reliable deviation detection in the ETL workflow.

E. Blockchain Integration for logging

Within the blockchain integration stage, all critical happenings within the ETL pipeline, from data extraction, transformation, and loading, are hashed and stored as log entries on the blockchain ledger, The metadata involved includes job start and end times, job IDs, cryptographic hashes of processed data, transformation types, and anomalies detected during processing. To conserve storage and protect privacy, the raw data itself never resides on the blockchain--only pertinent, non-sensitive details are logged. Smart contracts will regulate access to the environment based on agreed-upon rules while concurrently ensuring that ETL tasks occur in the correct order and report real-time inconsistencies. In this manner, the entire ETL life cycle becomes open, transparent, and tamper-proof, thereby considerably increasing information integrity, data security, and regulatory compliance.

Blockchain Platform Selection

Choosing the right blockchain platform is fundamental to developing a decentralized and secure ETL pipeline. Ethereum (via Ganache), Hyperledger Fabric, and Multichain are some hot choices, each presenting specific peculiarities that could serve particular organizational needs. Ethereum works well with Ganache in a stage-like environment for testing and development. It also supports smart contracts where the validation of ETL stages can be done automatically, and certain rules may be enforced from within the ETL process. The decentralized ledger of Ethereum makes sure that each transaction goes immortal, and in this way, it offers one layer of transparency and auditability. From this perspective, one can begin to consider Ethereum for any scenario that calls for public verifiability, a level of trustlessness, and an immutable log. Ganache especially allows developers to run a blockchain locally to

give quick feedback on iterations on the ETL logic and security model being developed.

On the other hand, Hyperledger Fabric and Multichain are better suited for private, enterprise-centric applications. Hyperledger Fabric is a permissioned blockchain framework offering modular components, such as pluggable consensus mechanisms and private data channels, making it highly configurable in the business environment. It is best suited for cases in which data privacy and access control take precedence, for example, while implementing sensitive applications in banking institutions or healthcare providers. Multichain is another private chain focused on ease of deployment, scalability, and management of access permissions. It allows users to handle large volumes of metadata securely without compromising on performance. Hyperledger Fabric and Multichain give much more control over the participant nodes and the visibility of transactions; this is fundamental to industries that are heavily regulated and governance-driven. In general, among these platforms, the choice should be driven by the specific needs of the ETL system, whether it calls for public transparency, private control, or fast prototyping.

ETL Pipeline Integration

The integration of blockchain technology into an ETL pipeline fundamentally alters the manner of recording, verifying, and securing data at every step of the operation: extraction, transformation, and loading. In the Extraction phase, data is pulled out from one or more sources such as transactional databases, sensors, APIs, or data lakes. The instant the extraction process starts, an integration with the blockchain-based system kicks off by logging important metadata: the identifier of the data source, the exact time of extraction, data type, and data volume. Every one of these details is cryptographically hashed using hashing algorithms such as SHA-256, giving rise to a unique name coined as the digital fingerprint of that extraction event. This hash is then written to the blockchain as an immutable transaction. As such, the system guarantees that the data's provenance and contextual clues are recorded in perpetuity, thereby eliminating any future disputes or unauthorized tampering. It also provides full traceability, which is important for audit purposes, compliance, and forensic investigations, setting a secure basis for later ETL operations.

During the Transform and Load phases, blockchain technology assures operational transparency and maintains a tamper-proof layer of security. Data undergoes some form of preprocessing, such as cleaning, filtering, normalization, encoding, or feature extraction, with each transformation step then recorded on the blockchain alongside certain metadata. Examples of such metadata include the transformation type, version of the algorithm, timestamp, and a hash of the transformed data. Should an inconsistency or anomaly arise in the dataset, the blockchain log could be used to correlate and trace the exact step and the responsible agent. Loaded into the destination, a Data Warehouse, an Analytics Engine, or Cloud-based Storage—the last operation is immutably recorded. The metadata could include a storage location identifier, load timestamp, status of the load job, and the hash of the final dataset. Continuous logging in all ETL phases builds an unbreakable, verifiable audit trail for operational diagnostics and regulatory UX compliance. Along the ETL fabric, by implementing blockchain, the system shifts pages from being traditionally dull to extremely transparent, secure, and accountable.

Smart Contract Enforcement

The most important aspect of smart contracts is the trust, security, and automation implemented into the blockchain-embedded ETL pipeline. Smart contracts, acting as programmable agents deployed over the blockchain, execute certain pre-defined rules autonomously and need no outside intervention. An important function of the contracts is access control, whereby they verify and authorize a user or system before an ETL job can commence. Therefore, only an approved role, such as a data engineer, system administrator, or certified automated script, can initiate or change processes listed under extraction, transformation, or loading. System-level permissions are embedded inside the contract; hence, access is forcibly enforced by every actor, ruling out the hazard of manual intervention. Other capabilities of these smart contracts are vested in ensuring compliance with the workflow logic of ETL operations. Thus, it confirms that each process is carried out in the pre-determined order-meaning, for example, that data cannot be loaded before transformation. Any violation of the defined workflow will result in either rejection or alert, thus preempting malicious, unauthorized operations or accidental missteps.

Besides, smart contracts provide for data integrity validation by cross-verifying cryptographic hashes generated at every stage of ETL. Should there be any mismatch or irregularity between the expected hash value and the actual hash value, it means a potential case of data tampering or corruption, which the contract instantly logs onto the blockchain immutably. When synergized with the MLP-GRU-based AI engine, an even more potent feature arises: in case the AI detects aberrant patterns such as irregular timing of execution or unexpected behaviour on the data, the smart contract can suspend operations autonomously, notify concerned parties, or reroute the job for further examination. Hence, this on-the-fly ETL interaction forms a self-regulating ETL system ensuring transparent and accountable data processing and intelligent reaction toward security threats. Embedding such governance directly in the pipeline fabric enables smart contracts to remove manual intervention, reduce latency in addressing threats, and increase compliance in sensitive, regulated environments.

Data Stored on the Blockchain

In the blockchain-embedded ETL framework, the data that is on the blockchain is carefully decided so that it adheres to transparency, security, and performance. Instead of raw data being stored on the chain, which would have

been expensive and computationally heavy, the blockchain hosts crucial metadata that provides traceability and accountability while keeping the systems efficient. The ETL job hash is one of the fundamental building blocks for the solution and serves as a cryptographic fingerprint (like SHA-256) that uniquely identifies the input, processing logic, and output of every single job. This makes it almost impossible to meddle with the system because any attempt to manipulate either the input, output, or process would result in a completely different hash. The system also logs timestamps of all ETL jobs, representing the moments when the jobs start and finish. They serve as a method for measuring pipeline latency and outage time and can therefore be correlated with security incidents or anomalous behavior.

The blockchain also saves job IDs, status flags (OK, error), summarization of execution, operational hashes, and timing data. These are all combined to give a complete picture of pipeline health and performance. Importantly, whenever the MLP-GRU-AI engine detects abnormalities from usual processing behavior, access from unauthorized users, or irregular throughput, anomaly flags are immediately raised and logged. The flags serve to enforce real-time monitoring and are finally stored in the audit trail used for forensic analysis and regulated reviews. Lastly, log hashes are recorded to maintain the integrity of detailed off-chain execution logs. Hence, auditors can validate the logs without touching sensitive internal information. By adopting a metadata-centric storage model, the blockchain implementation achieves availability for audits while maintaining inherent confidentiality and security, and verifiability of operations without compromising on efficiency or data privacy.

System Benefits and Reliability

In terms of ETL processes, the fused combination of blockchain and AI promises more transformative reliability, security, and transparency to systems. Fundamentally, blockchain imparts immutability to blockchain network records and stores every action taken on the ETL pipeline, whether it is data extraction, transformation, or loading, and these records cannot be forged or retroactively altered. It essentially confers an immutable audit trail for all involved parties to refer to for the accurate history of all operations recorded. Transparency within the network is maintained as job metadata, including timestamps, hashes, and status indicators, is publicly verifiable (for permissioned blockchain). This openness builds trust among various departments, external auditors, and regulatory bodies, which are empowered to trace the complete lineage of any data record without fear of manipulation or loss of integrity.

In effect, with an AI engine such as MLP-GRU, intelligent anomaly detections and response automation are derived. Any attempt at unauthorized access, process behaviors, or unauthorized job activity that causes anomalies in the system is immediately flagged, and the jobs may be suspended or escalated without manual intervention. Hence, opportunities for data breaches via malicious persons, insider threats, or accidental errors are reduced. Through blockchain-based logging, auditability becomes straightforward to comply with various stringent laws such as GDPR, HIPAA, and PCI-DSS. The auditors may retain immutable records for reviewing without requiring exhaustive manual tracking or documentation. This combination and this framework offer operational resilience, increased fault tolerance, and additional features demanded by the industries for a modern, secure, accountable, and automated data infrastructure in sensitive fields such as healthcare, finance, and telecommunication. This brings forth ETL systems of the future, which depart from just being technically sound and embrace the moral and legal responsibility of society.

The diagram for ETL on blockchain depicts a secure and intelligent data-processing pipeline, harnessing the power of blockchain technology and AI for ensuring data integrity, transparency, and timely threat mitigation, as depicted in Fig. 2. The steps begin with data extraction from any source such as a database, IoT device, or cloud repository, immediately followed by hashing of extracted metadata and logging it onto a blockchain. This ensures an immutable record of the origin and time of extraction of the data. During the transformation step, operations such as normalization or encoding occur, and all changes are recorded as hashed entries. Embedded smart contracts in the blockchain check every step of the ETL process, processing based on rules and verifying the sequence in which ETL steps are performed. If an anomaly or irregularity is found either in the job sequence or in the job itself, the AI modules powered by MLP-GRU raise an alert that is recorded immutably. In the loading step, the cleansed data is loaded into the destination warehouse or data lake, with additional metadata about job completion time, data volume, and transformation checksum being recorded.

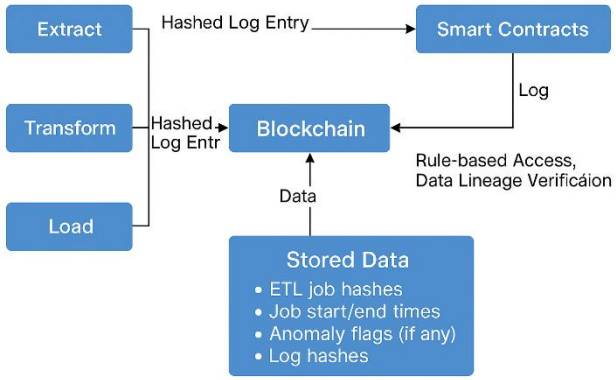

Fig. 2. Blockchain Framework

F. MLP-GRU for Anomaly Detection

Preprocessed ETL pipeline data are fed into a hybrid deep-learning model composed of an MLP and GRU for the detection of anomalies. It handles both anomaly detection and pattern characterization, learning static patterns with the

MLP architecture, while handling temporal dependencies with the GRU. Features like job duration, fluctuations in data volume, access frequency, hash mismatch indicators, and timestamp series are normalized and fed into the model. The GRU gains knowledge of both time-series trends and behavioral deviations after successive ETL executions, whereas the MLP compresses the dimensionality and classifies the feature-rich data. Hence, this setup helps spot any abnormally suspicious job executions, unusual delays, or security breach attempts. Anomalies detected are logged and flagged so that reactive measures and further analyses can be conducted. Thus, the step is vital in the proactive assessment and mitigation of risk in secure ETL systems.

Input Layer

The ETL system takes inputs in the form of features that are preprocessed in a particular manner, thereby ensuring that these features, spanning different operational metrics that vary temporally and structurally to detect anomalies, are well considered. Job duration measures how long an ETL task takes, where extraordinary values, either too high or too low, become suspicious for performance issues or for tampering or interference in the process. Hash mismatch indicators put forth the inconsistencies that come up between what is supposed to be the cryptographic hash and what comes about at a given stage of ETL-as-they-may-indicate a probable data manipulation or downright corruption. Another important feature is data volume changes, which could help detect abnormal rises or drops in the amount of data being processed, and such occurrences could be caused by intrusion, data loss, or even a system misconfiguration. By measuring these variations against historical baselines, the model will immediately flag a sudden change in behavior that is not in line with expected behavior.

Added to these are timestamp sequences that keep track of the temporal order and spacing of operations to catch timing anomalies such as unauthorized access during off-hours. Frequency of access denotes the number of times certain data sources or transformation scripts are activated, usually to recognize irregular returns or possible insider abuses. Transformation types are encoded to express the essence of operations (like normalization, aggregation, encoding) brought to the data, for different combinations of transformations may mark different sensitivity or risk. All these are normalized through Min-Max normalization, scaling all the features into a uniform range (say, 0 to 1), so that no feature potentially dominates any other and thus interferes with the whole learning process. This step is necessary for the stability and accuracy of the model, particularly when it faces static features such as job metadata in conjunction with dynamic features of time-series behaviors. This comprehensive, normalized set of features forms the input vector to the MLP-GRU model, thus permitting it to learn, detect, and respond to anomalies in real time.

MLP Component: Feature Abstraction

The MLP module has a crucial responsibility in abstracting and refining features coming from the ETL pipeline. As a feed-forward type of neural network, it takes normalized input features such as job duration, hash mismatches, transformation types, access frequencies, etc., passing these through one or two fully connected hidden layers equipped with activation functions like ReLU or Leaky ReLU. This ensures that the model is capable of learning nonlinear feature interactions. This stage acts primarily as the dimensionality reduction of inputs and brings into focus the most important aspects of data while eliminating irrelevant information that will act as noise. This process simplifies the feature space so that the next GRU layers would have a more relevant and distilled form of data representation.

In addition, dropout layers were embedded into the MLP architecture to alleviate overfitting and to increase the robustness of the model. During training, the model randomly disables a subset of units so that not too much importance can be assigned to any one feature, thereby providing a regularization mechanism to the network. MLP learning provides an additional layer of regularization to allow the model to generalize better on the unseen data, which is quite important in a setup where the ground truth is gathered for a relatively small set of anomalies that are either quite specialized or very subtle to ETL operations. Overall, the MLP segment acts as a pre-processing neural stage that transforms the raw feature vectors into an abstract low-dimensional embedding to be temporally considered by the GRU layers, enabling it to be an important module in the hybrid architecture for intelligent anomaly detection.

GRU Component: Temporal Pattern Learning

The GRU component models the temporal patterns and interdependencies existing in ETL pipeline installations. ETL operations tend to be time-based sequences, which include, for example, the sequence of operations, job run intervals, and access schedules. Using the GRU, these relations can be satisfactorily modeled. GRU has a gating mechanism standing in contrast to a classical RNN: it consists of an update and reset gate, which make decisions on how much of the past information should be kept or forgotten at a given time. This, in essence, holds more prolonged dependencies with no risk of gradient vanishing, allowing the model to notice subtle drifts in operational behavior over time. Additionally, the bidirectional GRU layers enable the model to factor in time in two directions, from both past and future time steps, to construct a more holistic timeline out of the ETL processes.

In practical terms, it signifies that the GRU can find slow-changing or context-dependent anomalies; for example, when a job becomes abnormal under certain conditions, following a series of prior events. This could mean the gradual increase in data volume and recurring hash mismatch patterns in consecutive ETL runs, which the GRU can almost perfectly learn and flag. Such time-sensitive

insight is needed to catch complex security threats or operational failures that cannot be spotted when they occur in isolated events, but through changes in behavior over time. The output of this GRU module, a sequence of hidden states that encode the temporal behavior learned by the GRU about the data pipeline, is then fed to the final anomaly scoring or classification layers. This makes the GRU an effective counterpart to the MLP layer while building a deep, intelligent system for secure real-time anomaly detection in ETL environments.

Output Layer

The output layer of this MLP-GRU architecture acts as a great final decision-making stage in the anomaly detection framework. Once the temporal sequence data has been processed and contextualized by the GRU part, the output is fed to a binary classification layer that uses either the sigmoid activation function to map the input to a probability value in the range [0, 1]. The value corresponds to the likelihood of the presence of an anomaly. Using some threshold (0.5 is common), the data are labeled normal (0) or anomalous (1). The softmax could instead be used wherever the output needs to be multi-class or have more detailed intermediate levels of certainty. This step transforms learned temporal and abstract patterns into a straightforward, interpretable label for immediate decision-making.

The output layer can also be defined to return confidence scores or severity scores for anomalies besides merely performing binary classification. These core values will allow us to assess how certain the model is about a particular prediction and how severe the anomaly may be. To illustrate, a high confidence score would imply that the system is quite certain about the occurrence of a real threat. In contrast, a low confidence score may indicate that it is merely suspicious or potentially risky behavior. These extra outputs are great for feeding into alerting systems or prioritizing responses in automated security workflows. Finally, the output layer takes all the complexity of multi-dimensional ETL activity patterns and converts them into actionable insights, thus enabling intervention well before any failure or breach.

Figure 3 exhibits the architecture of the proposed MLP-GRU model used for anomaly detection in blockchain-integrated secure ETL pipelines. The model can learn both spatial and temporal features of the data by combining the power of MLP and GRU. The leftmost side represents the input layer, which takes several features from the ETL process, such as data flow features, transformation logs, and access metadata. Such inputs are first fed into the MLP, where the hidden layer learns an abstract representation of the features for better classification. The MLP output is then fed into stacked GRU layers designed to capture temporal dependencies and sequential anomalies over ETL stages. With a gating mechanism, the GRU allows the model to keep or discard information selectively, hence improving the detection of subtle or evolving threats. Finally, the processed sequence is sent to an output layer for producing the prediction; the output is typically a classification of normal or anomalous behavior.

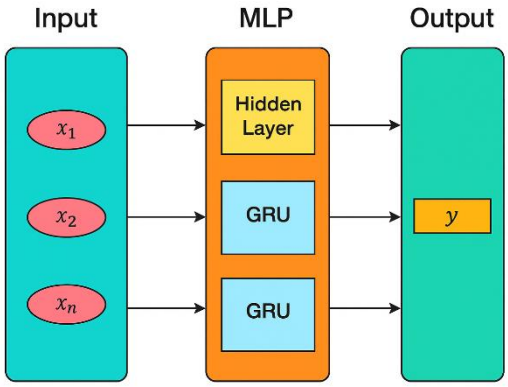

Fig 3. Architecture of MLP-GRU

IV. RESULTS & DISCUSSION

The results section provides a thorough evaluation of the proposed Blockchain-augmented ETL pipeline, techno-integrated with MLP-GRU anomaly detection. The experiments focused on addressing the system's performance on multiple fronts. Different visualizations in the form of bar charts and line graphs help prove the model's efficacy in anomaly identification and preservation of data integrity while processing data with the increasing volume of events at no compromise to performance. These findings have further demonstrated the hybrid framework's practicability and strength in bringing about secure, scalable, and intelligent ETL operations, particularly in data-sensitive and regulation-heavy fields.

A. Experimental Outcome

The latency-differentiating view of the three ETL pipeline architectures- Traditional ETL, Blockchain-Enhanced ETL, and Blockchain Integrated with AI ETL- from 100 to 10,000 record data volumes. As shown in Fig. 4, the Traditional ETL pipeline incurred the least latency in smaller data sizes, but its response time behaves very poorly when large numbers of data records are involved, indicating poor efficiency and limited scalability under heavy load situations. On the other hand, the Blockchain ETL pipeline shows ever so slightly higher latency at low volumes because of the cryptographic overhead associated with block creation, transaction validation, and ledger maintenance. However, its growth of latency remains rather controlled, presenting acceptable trade-offs in exchange for enhanced data integrity. The Blockchain + AI ETL framework achieves compromises and remains consistently lower concerning latency for Blockchain at higher volumes. The possibility is meant for AI to help with the intelligent scheduling of tasks, prediction of bottlenecks in processing, and intelligent routing of data within the pipeline.

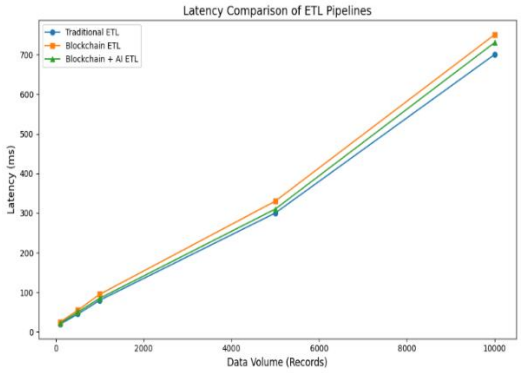

Fig 4. Latency

Throughput performance in Fig. 5, measured in records processed per second, has been illustrated for Traditional ETL, Blockchain-Enhanced ETL, and Blockchain + AI-based ETL systems against increasing data volumes. As the figure indicates, the Traditional ETL pipeline achieves the highest throughput at lesser data volumes, with minimal processing overhead on direct data handling; however, as the data volume increases, its throughput decreases drastically, making it an inefficient and unsuitable kind for heavy data analysis. The Blockchain ETL pipeline offers smaller throughput at the start, owing to the additional costs of cryptographic processing and consensus, with the performance cost accentuating as the data volume rises, further decreasing speed. Oddly enough, the Blockchain + AI ETL framework maintains a stable throughput curve through volumes of data. Even with a low-throughput solution when compared with Traditional ETL, the Blockchain + AI ETL framework can enforce intelligent task allocation, predictive optimization, and real-time anomaly detection, enabled by AI, to maintain high service efficiency adaptively.

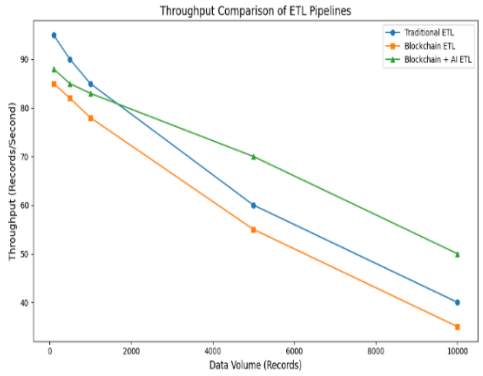

Fig 5. Throughput

Fig 6 presents the analytics of accuracy levels for anomaly detection embedded within three different types of ETL frameworks: Traditional ETL, Blockchain-Enhanced ETL, and Blockchain + AI, across the five test scenarios. Looking at the results, they serve to illustrate the highest levels of detection accuracy, forged by the Blockchain + AI ETL pipeline, which in all tests had an accuracy level of detection greater than 90%, setting the highest one at 95 percent in the last test. The reason for this high degree of accuracy is mostly due to the AI's side of learning from past experiences and patterns, from ever-flexible data behavior to raising alarms with fewer false positives. In comparison, the Blockchain-only ETL, with its feature of secure and immutable logging coupled with transparency in transactions to detect suspicious activities, had moderately higher accuracy than the traditional system. In real-time, however, it can not adapt intelligently to tweak detection. The Traditional ETL system, on the other hand, has the lowest and least stable accuracy, which keeps decreasing as test complexity rises-which is indicative of it being prone to undetected anomalies and being unable to automate threat responses.

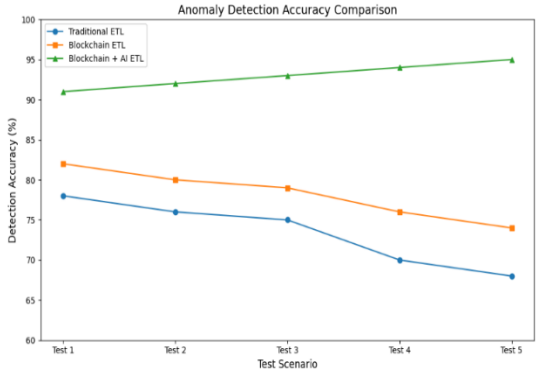

Fig. 6. Detection Accuracy

Figure 7 depicts the average block commit time concerning data volume for the two ETL system configurations, i.e., Blockchain ETL and Blockchain integrated with AI. As more data is processed, from 10 MB onwards to 1000 MB, in both systems, the commit time rises. Indeed, increasing the chunk of data processed entails higher computational efforts to marshal the data into blocks and to validate transactions in a blockchain network. Nonetheless, the Blockchain + AI method records lower commit times for all cases of data volume, indicating an efficient approach. Superiority in efficiency comes with AI choosing block sizes dynamically, limiting consensus contention, and anticipating peak loads to shorten transaction queue delays. Whereas the block commit time for Blockchain ETL quickly peaks at well beyond 25 seconds at the volume of about 1000 MB, the AI-based system arrests this upsurge to less than 20 seconds, thereby establishing higher scalability and responsiveness when high volume matters. In worth noting that Fig 7 thus very much emphasizes the practical value that the integration of AI might infuse into blockchain systems for scaling up the challenges of secure and large-scale ETL operations.

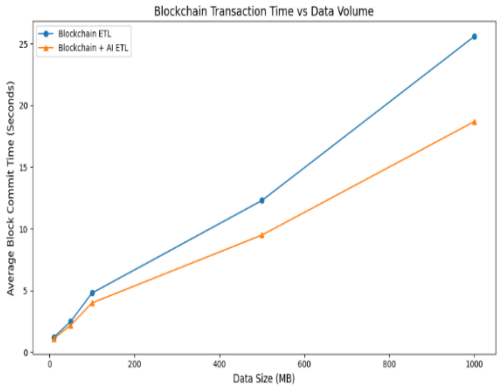

Fig 7. Transaction Time

Figure 8 illustrates a bar chart comprising key performance indicators for the proposed Blockchain + AI-based ETL-security method, presenting extremely high values for all evaluation measures. The model achieves an accuracy of 99%, which indicates the overall effectiveness in correctly identifying normal and anomalous ETL operations. The precision of 98.21% shows a strong ability to minimize false positives; the flagged anomalies are suspicious activities from a forensic perspective. A Recall rate of 98% makes this method stronger to ensure almost all relevant threats or anomalous events are considered without missing any critical ones. The F-1 score of 98.77%, a balanced value between precision and recall, proves the model's consistency and reliability in variable underpinning situations. That way, the results corroborate the usefulness of using integrated blockchain and AI for ETL-pipeline security, with threats being either detected with high accuracy or at least getting less attention from undetected ones or falsely flagged ones.

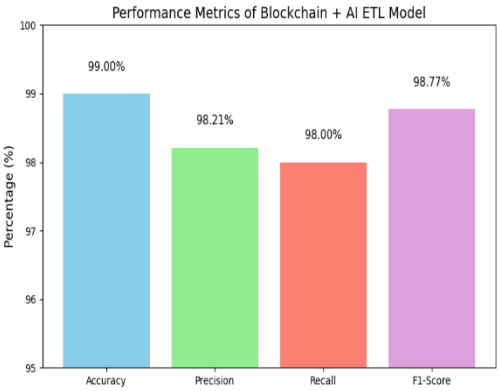

Fig 8. Performance Metrics

V. Conclusion and Future Work

This paper first advances the development of a secure ETL pipeline framework by interweaving Blockchain technology with an AI anomaly detection system that is based on MLP-GRU networks. A traditional ETL pipeline, even when standard encryption and firewall-based protection mechanisms are employed, remains subject to insider threats, tampering, and breaches of compliance owing to its centralized and opaque architecture. The hybrid model proposed in this paper aims to solve these issues by ensuring data immutability, traceability, and intelligent threat detection. In contrast, blockchain technology decentralizes the ledger, providing an auditable trail for every ETL operation, whereas the AI model based on MLP-GRU networks is capable of analyzing both static and sequential features of ETL logs to differentiate suspicious from legitimate operations in real-time. Experiments evaluate the performance of the model, effectively showing that it can detect threats while minimizing false alarms. Conversely, metrics such as block commit time and anomaly detection latency suggest that the hybrid architecture will continue to be scalable and responsive in the face of high data throughput, validating its suitability.

Several research paths lie ahead. Firstly, we could extend the system to support federated or edge ETL architectures where the AI models run locally near the data source and synchronize with the blockchain network. A second way to augment data privacy (while still maintaining verifiability) would be the integration of Zero-Knowledge Proofs (ZKPs) or Homomorphic Encryption with the blockchain. An additional research direction, of course, would be attempting to use more advanced AI models like Transformers or Graph Neural Networks (GNNs) for anomaly detection in relational or more complicated ETL settings. Last but not least, real-time dashboard integrations and automated regulatory audit generation using smart contracts could work toward better usability for an enterprise-level implementation. This research lays the foundation for a futuristic data engineering setting, secured, intelligent, and compliant by merging blockchain and AI along ETL pipelines.

References

[1] H. Matsuo *et al.*, "Diagnostic accuracy of deep-learning with anomaly detection for a small amount of imbalanced data: discriminating malignant parotid tumors in MRI," *Sci Rep*, vol. 10, no. 1, p. 19388, Nov. 2020, doi: 10.1038/s41598-020-76389-4.

[2] H. W. Oleiwi, D. N. Mhawi, and H. Al-Raweshidy, "MLTs-ADCNs: Machine Learning Techniques for Anomaly Detection in Communication Networks," *IEEE Access*, vol. 10, pp. 91006–91017, Aug. 2022, doi: 10.1109/ACCESS.2022.3201869.

[3] W. Marfo, D. K. Tosh, and S. V. Moore, "Network Anomaly Detection Using Federated Learning," in *MILCOM 2022 - 2022 IEEE Military Communications Conference (MILCOM)*, Rockville, MD, USA: IEEE, Nov. 2022, pp. 484–489. doi: 10.1109/MILCOM55135.2022.10017793.

[4] S. K. Vuppala, "LEVERAGING AI FOR PREDICTIVE ANALYTICS IN DATA SECURITY: IDENTIFYING AND PREVENTING ETL PIPELINE VULNERABILITIES," *IJAIRD*, vol. 3, no. 1, pp. 105–130, May 2025, doi: 10.34218/IJAIRD_03_01_008.

[5] M. Qasim and E. Verdu, "Video anomaly detection system using deep convolutional and recurrent models," *Results in Engineering*, vol. 18, p. 101026, Jun. 2023, doi: 10.1016/j.rineng.2023.101026.

[6] S. Mokhtari, A. Abbaspour, K. K. Yen, and A. Sargolzaei, "A Machine Learning Approach for Anomaly Detection in Industrial Control Systems Based on Measurement Data," *Electronics*, vol. 10, no. 4, p. 407, Feb. 2021, doi: 10.3390/electronics10040407.

[7] K. Al Jallad, M. Aljnidi, and M. S. Desouki, "Anomaly detection optimization using big data and deep learning to reduce false-positive," *J Big Data*, vol. 7, no. 1, p. 68, Dec. 2020, doi: 10.1186/s40537-020-00346-1.

[8] S. K. Vuppala, "AI-driven ETL Optimization for Security and Performance Tuning in Big Data Architectures," *IJLRP*.

[9] M. K. Hooshmand and D. Hosahalli, "Network anomaly detection using deep learning techniques," *CAAI Trans on Intel Tech*, vol. 7, no. 2, pp. 228–243, Jun. 2022, doi: 10.1049/cit2.12078.

[10] S. Kumar Vuppala, "ETL and BI in Military-Civilian Collaboration for Disaster Preparedness," *IJSR*, vol. 14, no. 5, pp. 639–649, May 2025, doi: 10.21275/SR25509004027.

[11] S. T. Ikram *et al.*, "Anomaly Detection Using XGBoost Ensemble of Deep Neural Network Models," *Cybernetics and Information Technologies*, vol. 21, no. 3, pp. 175–188, Sep. 2021, doi: 10.2478/cait-2021-0037.

[12] M. Said Elsayed, N.-A. Le-Khac, S. Dev, and A. D. Jurcut, "Network Anomaly Detection Using LSTM Based Autoencoder," in *Proceedings of the 16th ACM Symposium on QoS and Security for Wireless and Mobile Networks*, Alicante, Spain: ACM, Nov. 2020, pp. 37–45. doi: 10.1145/3416013.3426457.

[13] H. Son, Y. Jang, S.-E. Kim, D. Kim, and J.-W. Park, "Deep Learning-Based Anomaly Detection to Classify Inaccurate Data and Damaged Condition of a Cable-Stayed Bridge," *IEEE Access*, vol. 9, pp. 124549–124559, Jan. 2021, doi: 10.1109/ACCESS.2021.3100419.

[14] S. K. Vuppala, "SECURE AND COMPLIANT DATA MOVEMENT FOR SENSITIVE MILITARY OPERATIONS," *IJIS*, vol. 4, no. 1, pp. 39–68, Apr. 2025, doi: 10.34218/IJIS_04_01_003.

[15] P. Cichonski, T. Millar, T. Grance, and K. Scarfone, "Computer Security Incident Handling Guide : Recommendations of the National Institute of Standards and Technology," National Institute of Standards and Technology, NIST SP 800-61r2, Aug. 2023. doi: 10.6028/NIST.SP.800-61r2.

[16] R. Kumaran, "ETL Techniques for Structured and Unstructured Data," *SSRN Journal*, Jan. 2024, doi: 10.2139/ssrn.5143370.

[17] D. Seenivasan, "AI Driven Enhancement of ETL Workflows for Scalable and Efficient Cloud Data Engineering," *int. jour. eng. com. sci*, vol. 13, no. 06, pp. 26837–26848, Jun. 2024, doi: 10.18535/ijecs.v13i06.4824.

[18] M. F. Ansari, R. Sandilya, M. Javed, and D. Doermann, "ETLNet: An Efficient TCN-BiLSTM Network for Road Anomaly Detection Using Smartphone Sensors," Jun. 2024, *arXiv*. doi: 10.48550/ARXIV.2412.04990.

[19] N. Joshi, "Optimizing Real-Time ETL Pipelines Using Machine Learning Techniques," Aug. 2024, *SSRN*. doi: 10.2139/ssrn.5054767.

[20] Saswata Dey, Writuraj Sarma, and Sundar Tiwari, "Deep learning applications for real-time cybersecurity threat analysis in distributed cloud systems," *World J. Adv. Res. Rev.*, vol. 17, no. 3, pp. 1044–1058, Mar. 2023, doi: 10.30574/wjarr.2023.17.3.0288.

[21] S. Hiremath *et al.*, "A New Approach to Data Analysis Using Machine Learning for Cybersecurity," *BDCC*, vol. 7, no. 4, p. 176, Nov. 2023, doi: 10.3390/bdcc7040176.

[22] O. Hamza, A. Collins, A. Eweje, and G. O. Babatunde, "Advancing Data Migration and Virtualization Techniques: ETL-Driven Strategies for Oracle BI and Salesforce Integration in Agile Environments," *IJMRGE*, vol. 5, no. 1, pp. 1100–1118, Jan. 2024, doi: 10.54660/.IJMRGE.2024.5.1.1100-1118.

[23] S. S. Aljameel *et al.*, "An Anomaly Detection Model for Oil and Gas Pipelines Using Machine Learning," *Computation*, vol. 10, no. 8, p. 138, Aug. 2022, doi: 10.3390/computation10080138.

[24] S. Akcay, D. Ameln, A. Vaidya, B. Lakshmanan, N. Ahuja, and U. Genc, "Anomalib: A Deep Learning Library for Anomaly Detection," in *2022 IEEE International Conference on Image Processing (ICIP)*, Bordeaux, France: IEEE, Oct. 2022, pp. 1706–1710. doi: 10.1109/ICIP46576.2022.9897283.