

Jayendra S. Jadhav  
Jyoti Deshmukh

Research paper  
DOI -- 10.24874/QF.25.054



## A ROBUST BLOCKCHAIN-BASED FRAMEWORK FOR MANAGING UNKNOWN VIRAL DISEASES IN HEALTHCARE SUPPLY CHAINS USING MACHINE LEARNING

**Abstract:** *Objective of the Study: Frequent viral outbreaks demand a secure, scalable system for early disease detection and healthcare supply chain management. This study proposes a Blockchain-integrated Machine Learning framework to enhance disease surveillance, ensure secure data exchange, and optimize healthcare logistics.*

*Methodology/Approach: A multi-layered architecture integrates Blockchain for decentralized data management and Machine Learning for symptom clustering, anomaly detection, and predictive analytics. Smart contracts validate data, while cryptographic techniques ensure privacy. Geospatial analysis maps real-time symptom trends and medicine demand fluctuations.*

*Originality/Relevance: Unlike centralized systems prone to inefficiencies and data breaches, this framework offers a transparent and decentralized solution. By combining predictive analytics with Blockchain, it strengthens disease tracking, outbreak preparedness, and resource allocation.*

*Main Results: The system improves disease detection accuracy, enhances supply chain transparency, and ensures secure data exchange. Machine Learning models accurately identify emerging disease patterns, while Blockchain guarantees data immutability. Geospatial analytics predict medicine shortages and outbreak hotspots.*

*Theoretical/Methodological Contributions: This research enhances Blockchain-based healthcare models with privacy-preserving cryptographic mechanisms and federated Machine Learning. It ensures interoperability among healthcare providers and regulatory bodies while promoting secure data sharing.*

*Social/Management Contributions: By improving pandemic preparedness and healthcare supply chain resilience, this framework ensures equitable medical resource distribution, faster containment, and better public health security. Its transparency builds trust, aiding data-driven policy decisions and minimizing medicine shortages.*

**Keywords:** *Blockchain Technology, Machine Learning, Early Disease Detection, Viral Disease*

<sup>1</sup> Corresponding author: Jayendra S. Jadhav  
Email: [jayendra071985@gmail.com](mailto:jayendra071985@gmail.com)

## 1. Introduction

The recent global health crises have exposed significant weaknesses in healthcare systems, particularly in their ability to manage and respond effectively to novel viral diseases. This research introduces a ground-breaking Blockchain-based framework designed to strengthen the oversight of viral disease outbreaks within the healthcare supply chain. By combining advanced mechanisms for early detection, robust systems for safeguarding patient data, and agile strategies for supply chain management, this approach seeks to enhance the preparedness and adaptability of healthcare services significantly (Jadhav & Deshmukh, 2024).

### 1.1. Early Detection of Unprecedented Viral Diseases

Timely identification of viral diseases is crucial to curbing their potential to cause extensive outbreaks. This strategy focuses on utilizing early symptoms reported by patients to enable the prompt recognition of novel viral illnesses. By monitoring and analysing symptom trends within a decentralized network, healthcare professionals can swiftly detect and address emerging viral threats before they escalate to critical stages. This proactive monitoring framework empowers both patients and healthcare providers to respond effectively, enhancing readiness and minimizing the impact of outbreaks (Jadhav & Deshmukh, 2024).

### 1.2. Blockchain for Enhanced Record Keeping and Supply Chain Oversight

Blockchain technology offers significant dual benefits in healthcare. Firstly, it establishes a highly secure infrastructure for managing Electronic Health Records (EHRs), safeguarding patient information against unauthorized access and alterations. Secondly, it functions as a foundational element in optimizing medical supply chain

management. By enabling the creation of a transparent and unalterable ledger, Blockchain fosters trust among all stakeholders—patients, healthcare providers, pharmacies, distributors, manufacturers, and suppliers. This capability is especially critical during high-demand situations like pandemics, where reliability and coordination are paramount (Jadhav & Deshmukh, 2022).

### 1.3. Machine Learning Analysis for Disease Detection

Machine learning serves as a cornerstone of the proposed system by focusing on two essential dimensions for analysis:

- **Patient Symptom Trends:** By monitoring the frequency and correlation of common symptoms across populations, machine learning algorithms can predict the emergence of unknown viral diseases, enabling pre-emptive healthcare responses (Jadhav & Deshmukh, 2024).
- **Medicine Demand Surges:** Simultaneously, these algorithms assess changes in the demand for specific medications and drugs, providing early indicators of a potential outbreak. This data is vital for managing stock levels and distribution logistics, ensuring that essential medicines are available where they are most needed (Ibrahim et al., 2022).

### 1.4. Integrating Stakeholders in Medical Supply Chain Management

Effective management of the medical supply chain is pivotal in combating viral diseases. Coordination among key stakeholders—patients, doctors, pharmacies, distributors, manufacturers, and suppliers—is essential. A Blockchain-powered platform guarantees that all stakeholders can seamlessly access up-to-date information about inventory

levels, shipping progress, and delivery timelines in real time. This transparency helps in maintaining a steady flow of supplies, preventing bottlenecks, and facilitating swift responses to emergent needs during a health crisis (Vora et al., 2023).

The proposed Blockchain-based system addresses critical gaps in the management of healthcare supply chains during viral disease outbreaks. By integrating early symptom detection with robust supply chain management and predictive analytics through machine learning, this approach not only enhances disease surveillance but also improves the overall responsiveness of healthcare systems. This innovative solution is a step forward in preparing for and managing future pandemics, ensuring a resilient healthcare infrastructure.

## 2. Literature survey

The integration of Blockchain technology and machine learning in healthcare has been the subject of extensive research, particularly in response to the growing challenges of managing viral diseases. This literature survey examines research focused on early disease detection, data security, efficient supply chain management, and stakeholder collaboration within healthcare systems.

### 2.1. Understanding Categories and Subcategories of Viral Diseases

- **Respiratory Viral Infections:** This category includes illnesses that primarily affect the respiratory tract, such as:
  - *Common Cold:* Symptoms include a runny nose, cough, and mild fever (Centers for Disease Control and Prevention, 2023).
  - *Flu (Influenza):* Marked by high fever, muscle aches, and

fatigue (Centers for Disease Control and Prevention, 2023).

- *COVID-19:* Symptoms range from cough and shortness of breath to loss of taste and smell (World Health Organization, 2023).

- **Exanthematous Viral Infections:** These diseases are characterized by skin rashes. Subcategories include:

- *Measles:* Rashes that start on the face and spread across the body, often accompanied by fever (Banerjee et al., 2020).
- *Chickenpox:* Itchy, fluid-filled blisters that appear in clusters (Johnson, 2023).
- *Rubella:* A mild rash, often starting on the face and spreading downward (World Health Organization, 2023).

- **Gastrointestinal Viral Infections:** These infections primarily affect the digestive system and include:

- *Norovirus:* Causes sudden, severe diarrhea and vomiting, typically in adults (Centers for Disease Control and Prevention, 2023).
- *Rotavirus:* Common in children, leading to dehydration from diarrhea and vomiting (Glass et al., 2017).

- **Systemic Viral Infections:** Systemic infections impact multiple body systems. Examples include:

- *Dengue Fever:* High fever, severe body aches, and potential complications like a drop in platelet count (World Health Organization, 2023).

- **Localized Infections:** These infections affect specific areas of the body:

- *Genital Viral Infections:* The condition, resulting from the Herpes Simplex Virus (HSV), manifests as painful sores or

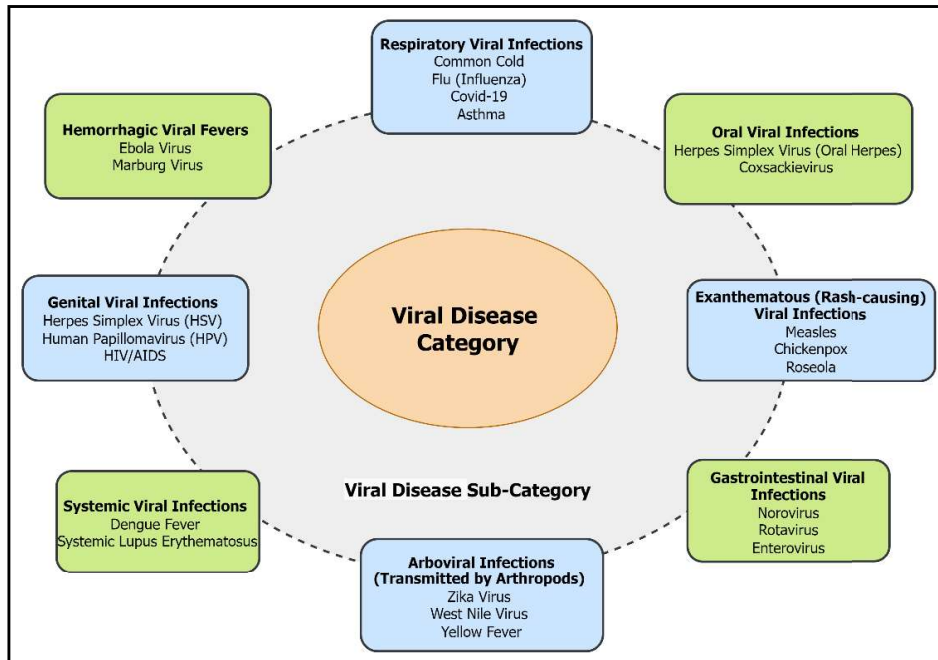
blisters in the genital region (Farley et al., 2020).

- *Oral Viral Infections*: HSV may also cause cold sores or blisters around the mouth (Banerjee et al., 2020).
- **Arboviral Infections**: Transmitted by arthropods (like mosquitoes), this category includes:
  - *Zika Virus*: Known for neurological complications and, in some cases, birth defects (World Health Organization, 2023).
  - *Yellow Fever*: Causes fever, jaundice, and potential organ

damage (Centers for Disease Control and Prevention, 2023).

- *Chikungunya*: Similar symptoms to dengue but often includes persistent joint pain (Centers for Disease Control and Prevention, 2023).

- **Hemorrhagic Viral Fevers**: These severe illnesses are characterized by bleeding disorders and systemic symptoms:
  - *Ebola and Marburg Viruses*: Lead to symptoms like fever, severe bleeding, and organ failure (Feldmann & Geisbert, 2011).



**Figure 1.** Viral Disease Category and Subcategory

Identifying symptoms is crucial for accurately diagnosing viral diseases, as each category and subcategory has distinct features. For example, respiratory infections like the common cold and flu differ in severity, while Exanthematous infections such as measles and chickenpox vary in rash type and progression. Gastrointestinal infections like norovirus cause sudden,

severe symptoms, whereas rotavirus is more common in children. Systemic infections, including dengue and chikungunya, overlap but differ in rash patterns and joint pain severity. Similarly, Arboviral and Hemorrhagic fevers, such as Zika and Ebola, are distinguished by neurological effects versus severe bleeding disorders.

## 2.2. Early Detection of Viral Diseases

Early identification of new viral diseases is crucial to prevent widespread outbreaks. Leveraging technologies like Blockchain and machine learning has proven effective for early detection and response.

- Patient Symptom Surveillance:** Combining machine learning and Blockchain technology can significantly enhance early detection systems. Jadhav and Deshmukh (Jadhav & Deshmukh, 2024) emphasize the role of these technologies in addressing challenges during COVID-19 by providing secure, decentralized data analysis for early symptom detection. This approach helps to promptly recognize patterns in patient-reported symptoms. Furthermore, Gupta et al. (Gupta et al., 2023) explored how Blockchain could secure IoT-driven healthcare systems to

analyse real-time patient data securely, facilitating timely identification of emerging threats (Jadhav & Deshmukh, 2025).

- Data Aggregation for Trend Detection:** Aggregating and analysing patient data effectively is essential for accurate predictions. Blockchain's decentralized approach ensures privacy and data integrity while allowing large-scale data sharing, as outlined by Kumar et al. (Kumar et al., 2024). This data security promotes more reliable and timely detection of viral trends. Radanovic and Likić (Radanović & Likić, 2018) also identify opportunities for Blockchain in securely collecting health data to aid public health monitoring.

The table 1 provides a summary of the section.

**Table. 1** Section 2.1 Summary

| Reference Number          | Paper Description   | Analysis   |
|---------------------------|---|--|
| (Jadhav & Deshmukh, 2024) | Investigates the use of machine learning combined with Blockchain technology to enable early detection of diseases, with a particular emphasis on COVID-19. | <b>Pros:</b> Enhances early detection; secures patient data<br><b>Cons:</b> Limited to specific cases (COVID-19); scalability issues<br><b>Technology:</b> ML, Blockchain                                  |
| (Gupta et al., 2023)      | Discusses Blockchain's use for securing IoT-driven healthcare data and real-time analysis.  | <b>Pros:</b> Improves data security; enables real-time analysis<br><b>Cons:</b> Complex integration with existing systems, Scalability not discussed<br><b>Technology:</b> IoT, Blockchain                 |
| (Kumar et al., 2024)      | Offers a comprehensive analysis of the role of Blockchain in healthcare, particularly its application in data analysis for early disease detection systems. | <b>Pros:</b> Provides broad perspective on data use, Focuses on large-scale data analysis<br><b>Cons:</b> Lacks practical application details, No focus on specific cases<br><b>Technology:</b> Blockchain |
| (Jadhav & Deshmukh, 2025) | Highlights Blockchain's potential in gathering and sharing patient data securely for outbreak.  | <b>Pros:</b> Highlights data integrity benefits, Emphasizes data security<br><b>Cons:</b> Does not address large-scale implementation,<br><b>Technology:</b> Blockchain                                    |

## 2.3. Blockchain for Record-Keeping and Supply Chain Oversight

Blockchain has a transformative role in

securely managing healthcare records and improving the supply chain. It enhances transparency and efficiency in record-keeping and logistics, which are critical

during health crises.

*Securing Electronic Health Records (EHRs)*: Ensuring the security of Electronic Health Records (EHRs) is critical for safeguarding patient data from unauthorized access and manipulation.

Blockchain technology plays a key role in providing a secure storage solution for EHRs within the healthcare supply chain, as highlighted by Jadhav and Deshmukh (Jadhav & Deshmukh, 2022). Additionally, Liu et al. (Liu et al., 2020) propose an advanced approach by integrating a user revocation mechanism within Ciphertext-Policy Attribute-Based Encryption (CP-ABE) systems, further enhancing the protection of patient data with improved privacy and access control measures.

*Transparent and Efficient Supply Chains*: Transparency in medical supply chains is essential to prevent counterfeiting and ensure timely delivery of healthcare products.

Blockchain creates a tamper-proof ledger for tracking pharmaceuticals across their journey, enhancing transparency. Vora et al. (2023) highlight how this improves overall supply chain efficiency, reducing errors and delays. Similarly, Ali et al. (Ali et al., 2021) describe how Blockchain enables real-time data-sharing among stakeholders for better oversight.

*Automating Inventory Through Smart Contracts*: Blockchain-based smart contracts automate key processes in supply chains, such as inventory restocking. Liu et al. (Liu et al., 2022) explore how smart contracts ensure a continuous supply of medicines by triggering orders automatically when stock levels fall below a threshold, thus addressing sudden surges in demand efficiently.

The table 2 provides a summary of the section.

**Table. 2** Section 2.2 Summary

| Reference Number          | Paper Description   | Analysis   |
|---------------------------|---|--|
| (Jadhav & Deshmukh, 2022) | Reviews Blockchain's application in healthcare supply chains with a focus on securing Electronic Health Records (EHRs).                                   | <ul style="list-style-type: none"> <li>• <b>Pros</b>: Ensures EHR security; improves data integrity</li> <li>• <b>Cons</b>: Challenges in real-world supply chain implementation</li> <li>• <b>Technology</b>: Blockchain</li> </ul>                             |
| (Vora et al., 2023)       | Introduces a Blockchain-driven framework for managing the medical supply chain, designed to combat the issue of counterfeiting.                           | <ul style="list-style-type: none"> <li>• <b>Pros</b>: Enhances transparency in supply chains; Prevents counterfeit drugs.</li> <li>• <b>Cons</b>: High setup costs; Technical complexities</li> <li>• <b>Technology</b>: Blockchain.</li> </ul>                  |
| (Liu et al., 2022)        | Analyses the use of smart contracts in automating healthcare supply chain transactions.   | <ul style="list-style-type: none"> <li>• <b>Pros</b>: Automates supply chain processes; Reduces human errors</li> <li>• <b>Cons</b>: Technical deployment barriers</li> <li>• <b>Technology</b>: Smart Contracts, Blockchain</li> </ul>                          |
| (Ali et al., 2021)        | Examines Blockchain-based solutions for self-sovereign patient identity management in healthcare settings.  | <ul style="list-style-type: none"> <li>• <b>Pros</b>: Enhances patient data control; Improves data privacy</li> <li>• <b>Cons</b>: Potential regulatory and interoperability issues</li> <li>• <b>Technology</b>: Self-Sovereign Identity, Blockchain</li> </ul> |
| (Liu et al., 2020)        | Suggests the implementation of a user revocation framework within Ciphertext-Policy Attribute-Based Encryption (CP-ABE) systems to enhance data security. | <ul style="list-style-type: none"> <li>• <b>Pros</b>: Enhances data security; Enables user access control</li> <li>• <b>Cons</b>: Complexity in implementation</li> <li>• <b>Technology</b>: CP-ABE, Blockchain</li> </ul>                                       |

## 2.4. Machine Learning for Disease Detection

Machine learning is increasingly used to analyse health data, providing timely insights that improve resource management and disease response.

- **Predictive Analytics for Symptom Trends:** Machine learning algorithms help detect emerging diseases by analysing patient data trends on Blockchain-secured networks. Jadhav and Deshmukh (Jadhav & Deshmukh, 2024) emphasize that using machine learning can efficiently analyse large-scale health data, enabling early detection. Studies by Ibrahim et al. (Ibrahim et al., 2022) and Naruse et al. (Naruse et al., 2022) support this approach, demonstrating how machine learning can predict and recognize unusual symptom patterns for prompt healthcare response.
- **Medicine Demand Forecasting:** Accurate forecasting of medicine demand is crucial during outbreaks

to manage supply chains effectively. Salah et al. (Salah et al., 2019) discuss how integrating machine learning with Blockchain systems can forecast medicine demand and prevent shortages. Similarly, the work by Gupta and Malhotra (Gupta & Malhotra, 2021) shows that analysing real-time data on medicine usage within Blockchain networks allows stakeholders to optimize drug distribution.

- **Automated Diagnosis Tools:** Machine learning can also enhance diagnosis by rapidly processing patient-reported symptoms to detect diseases. Zhang et al. (Zhang et al., 2021) and Ali et al. (Ali et al., 2021) discuss how automated diagnosis systems, combined with Blockchain-secured data, ensure accuracy while maintaining data confidentiality, helping to quickly identify emerging health threats.

The table 3 provides a summary of the section.

**Table 3.** Section 2.3 Summary

| Reference Number          | Paper Description   | Analysis  |
|---------------------------|---|---|
| (Jadhav & Deshmukh, 2024) | Explores machine learning models for analysing patient-reported symptoms on a Blockchain-secured network. | <ul style="list-style-type: none"> <li>• <b>Pros:</b> Enhances early detection; Secures patient data</li> <li>• <b>Cons:</b> Limited to COVID-19; Scalability not explored</li> <li>• <b>Technology:</b> Machine Learning, Blockchain</li> </ul>        |
| (Ibrahim et al., 2022)    | Highlights the use of machine learning in predicting medicine demand within healthcare supply chains.     | <ul style="list-style-type: none"> <li>• <b>Pros:</b> Improves inventory management; Aids logistics planning</li> <li>• <b>Cons:</b> Requires large datasets; Potential privacy concerns</li> <li>• <b>Technology:</b> Machine Learning</li> </ul>      |
| (Naruse et al., 2022)     | Proposes attribute-based encryption with Blockchain for data security in healthcare systems.              | <ul style="list-style-type: none"> <li>• <b>Pros:</b> Supports data privacy; Strengthens encryption and security</li> <li>• <b>Cons:</b> High computational overhead</li> <li>• <b>Technology:</b> Attribute-Based Encryption, Blockchain</li> </ul>    |
| (Naruse et al., 2022)     | Reviews Blockchain's collaboration with AI for predicting medicine demand surges.                         | <ul style="list-style-type: none"> <li>• <b>Pros:</b> Identifies AI integration in Blockchain; Highlights potential research areas</li> <li>• <b>Cons:</b> Lacks practical application examples</li> <li>• <b>Technology:</b> AI, Blockchain</li> </ul> |
| (Gupta & Malhotra, 2021)  | Explores Blockchain integration with IoT for managing real-time healthcare data.                          | <ul style="list-style-type: none"> <li>• <b>Pros:</b> Improves real-time data analysis; Enhances security in data management</li> <li>• <b>Cons:</b> Complex integration with existing systems</li> <li>• <b>Technology:</b> IoT, Blockchain</li> </ul> |

## 2.5. Integrating Stakeholders in Supply Chain Management

A cohesive approach involving all stakeholders—patients, healthcare providers, pharmacies, distributors, and suppliers—is essential for efficient medical supply chain management.

- **Blockchain-Enabled Collaboration:** A Blockchain-enabled system allows real-time, transparent communication among all stakeholders. Jadhav and Deshmukh (Jadhav & Deshmukh, 2022) explain how such systems enable data sharing, helping parties maintain awareness of stock levels, shipments, and supply needs. This enhanced visibility facilitates quicker responses to emerging needs. Vora et al. (2023) also discuss Blockchain's potential to improve coordination among stakeholders, ensuring smooth and efficient supply chain operations.
- **Optimizing Supply Chain Logistics:** Albassam et al. (Albassam et al., 2024) and Wen et al. (Wen et al.,

2020) provide frameworks for using Blockchain to optimize supply chains, making it easier to track medical supplies from the manufacturing stage to final delivery. Real-time data access and secure, tamper-proof records reduce delays and ensure timely distribution of essential items.

- **Real-Time Tracking and Traceability:** Blockchain technology plays a crucial role in ensuring full traceability throughout medical supply chains, which is essential for efficient distribution of resources. Houtan et al. (Houtan et al., 2020) suggest the implementation of a self-sovereign identity system to uphold the authenticity of products, reducing the chances of counterfeit drugs entering the healthcare supply chain. This approach ensures the safety and dependability of the medical supplies being distributed.

The table 4 provides a summary of the section.

**Table 4.** Section 2.4 Summary

| Reference Number        | Paper Description   | Analysis   |
|-------------------------|---|--|
| (Albassam et al., 2024) | Explores the use of Blockchain technology combined with cloud computing to enhance supply chain management within telemedicine systems. | <ul style="list-style-type: none"> <li>• <b>Pros:</b> Enhances data security; Promotes seamless telemedicine</li> <li>• <b>Cons:</b> Scalability in large systems remains a challenge</li> <li>• <b>Technology:</b> Blockchain, Cloud Computing</li> </ul> |
| (Wen et al., 2020)      | Analyses Blockchain's role in providing real-time tracking and traceability in healthcare supply chains.                                | <ul style="list-style-type: none"> <li>• <b>Pros:</b> Addresses data privacy and security; Enhances data traceability</li> <li>• <b>Cons:</b> Implementation challenges in varied ecosystems</li> <li>• <b>Technology:</b> Blockchain</li> </ul>           |
| (Houtan et al., 2020)   | Explores the use of Blockchain for self-sovereign patient identities, aiming to improve the traceability of healthcare logistics.       | <ul style="list-style-type: none"> <li>• <b>Pros:</b> Enhances patient data ownership; Supports data traceability</li> <li>• <b>Cons:</b> Requires widespread adoption</li> <li>• <b>Technology:</b></li> </ul>  |

Recent studies highlight the revolutionary capabilities of Blockchain and machine learning in the healthcare sector. Blockchain provides a secure, transparent method for managing data, while machine learning

leverages predictive analytics to enhance decision-making processes. Together, these technologies create a robust framework that improves early detection of diseases, streamlines supply chains, and fosters better

cooperation among various stakeholders. This combined approach not only solves present-day challenges but also paves the way for developing a more resilient healthcare system to better handle future health emergencies.

## **2.6. Findings and discussion**

The combination of Blockchain technology and machine learning holds great promise in tackling the challenges associated with managing viral diseases. Studies emphasize the importance of early detection in preventing large-scale outbreaks, and the integration of these technologies can significantly improve this process. For instance, research by Jadhav and Deshmukh (Jadhav & Deshmukh, 2024) illustrates how machine learning models, when combined with Blockchain, can effectively analyze symptoms reported by patients, aiding in the early identification of potential health threats. Similarly, Gupta et al. (Gupta et al., 2023) highlighted the critical role of Blockchain in ensuring secure real-time processing of patient data within IoT-based healthcare systems, thus facilitating the prompt detection of emerging viral risks.

Blockchain also plays a transformative role in secure record-keeping and supply chain oversight. By creating a transparent, tamper-proof ledger, it ensures data integrity and patient privacy. Jadhav and Deshmukh (Jadhav & Deshmukh, 2022) emphasized the importance of using Blockchain to secure Electronic Health Records (EHRs), protecting patient data from unauthorized access. In the context of supply chain management, Vora et al. (2023) discussed Blockchain's potential to prevent counterfeiting by tracking pharmaceuticals and ensuring timely delivery. This transparent tracking system, along with the use of smart contracts to automate inventory management (Liu et al., 2022), enables healthcare providers to maintain a continuous supply of essential medicines, enhancing overall healthcare efficiency.

Machine learning further contributes to healthcare optimization by predicting disease trends and managing resource allocation. According to Jadhav and Deshmukh (Jadhav & Deshmukh, 2024), machine learning algorithms can analyse patient symptom data on Blockchain-secured networks, allowing for early intervention in emerging diseases. Moreover, predictive models, as described by Ibrahim et al. (Ibrahim et al., 2022), help forecast medicine demand during outbreaks, optimizing inventory management and logistics. Studies such as those by Salah et al. (Salah et al., 2019) indicate that integrating AI with Blockchain enables better management of healthcare supplies, preventing shortages and ensuring patient access to necessary medications.

Finally, the literature stresses the importance of stakeholder integration in the healthcare supply chain. Blockchain facilitates real-time communication among patients, healthcare providers, pharmacies, and suppliers, streamlining logistics. Research by Albassam et al. (Albassam et al., 2024) and Wen et al. (Wen et al., 2020) shows that Blockchain-enabled collaboration enhances data security, transparency, and traceability in the supply chain, while reducing delays in distributing essential supplies.

In summary, these studies collectively highlight how Blockchain and machine learning can revolutionize healthcare management, offering solutions for early disease detection, secure data handling, predictive analytics, and supply chain optimization. However, despite these advantages, certain challenges and gaps in research remain.

## **2.7. Challenges in Implementing Blockchain for Identifying Unknown Diseases**

Implementing Blockchain technology for identifying unknown diseases presents several challenges that need to be addressed to ensure its effectiveness and efficiency. The key barriers to adopting Blockchain for

disease identification include the complexities of handling sensitive medical data, ensuring seamless integration with existing healthcare systems, and overcoming technological and legal challenges to facilitate the accurate and timely identification of unknown diseases. This section delves into each issue, highlighting existing challenges and their influence.

#### **Data Collection and Validation:**

Identifying unknown diseases relies on gathering extensive data, including patient symptoms, genetics, and environmental factors. However, ensuring this data's accuracy and integrity is challenging.

- *Challenge:* Blockchain offers immutable storage, but it does not inherently verify data accuracy. Jadhav and Deshmukh (Jadhav & Deshmukh, 2024) point out the difficulty of maintaining data integrity in decentralized settings. Gupta et al. (Gupta et al., 2023) further highlight the challenge of validating real-time IoT healthcare data before storage on the Blockchain.
- *Influence:* Without reliable validation mechanisms, Blockchain systems risk storing inaccurate data, leading to potential misdiagnoses or false identification of diseases.

**Data Standardization and Interoperability:** The variety of data formats used by healthcare providers makes seamless data exchange difficult.

- *Challenge:* Standardized protocols for cross-platform data exchange are lacking. Kumar et al. (Kumar et al., 2024) note that current Blockchain implementations struggle to manage diverse data formats, limiting large-scale data aggregation for disease identification.
- *Influence:* This lack of interoperability reduces the effectiveness of Blockchain in consolidating data from various sources, impacting the quality of early detection analytics.

#### **Privacy-Preserving Data Sharing:**

Detecting unknown diseases often requires sharing sensitive patient data while ensuring privacy.

- *Challenge:* Techniques like zero-knowledge proofs and homomorphic encryption offer potential solutions, but their practical implementation in real-time Blockchain systems is not yet fully explored (Liu et al., 2020). Liu et al. (Liu et al., 2020) proposed a user revocation scheme using Ciphertext-Policy Attribute-Based Encryption (CP-ABE), yet integrating such privacy controls into Blockchain networks remains a challenge.
- *Influence:* Incorporating these privacy-preserving methods is critical for compliance with regulations like HIPAA and GDPR but could introduce additional computational costs (Liu et al., 2020), (Gupta & Malhotra, 2021).

**Scalability and Data Storage:** Effective disease identification requires analysing large datasets, which strains Blockchain networks.

- *Challenge:* Current Blockchain systems face storage capacity limitations and high computational demands. Ali et al. (Ali et al., 2021) and Houtan et al. (Houtan et al., 2020) discuss the constraints of Blockchain in managing large-scale health data. Research into off-chain storage, like IPFS integration, is needed to support extensive data management.
- *Influence:* Without addressing scalability, Blockchain systems may fail to facilitate the real-time analysis required for disease identification.

#### **Consensus Mechanisms for Medical Data:**

Traditional Blockchain consensus mechanisms like Proof of Work (PoW) are not suited for healthcare due to high energy consumption and latency.

- *Challenge:* Existing protocols are not optimized for processing sensitive

medical data in real-time. Radanovic and Likic (Radanović & Likić, 2018) suggest developing fast, secure, and energy-efficient consensus algorithms tailored to healthcare data processing.

- *Influence:* Tailored consensus mechanisms could significantly enhance Blockchain's capacity for accurate and timely data analysis in disease identification.

**Integration of Machine Learning and Blockchain:** Machine learning models require access to large datasets for training, posing challenges for secure data access on Blockchain.

- *Challenge:* Jadhav and Deshmukh (Jadhav & Deshmukh, 2024) identify difficulties in preserving data privacy during the machine learning process. Privacy-preserving techniques, like federated learning on Blockchain networks, need further exploration to maintain data integrity (Jadhav & Deshmukh, 2024).
- *Influence:* Creating a robust framework for integrating machine learning with Blockchain would enable more precise disease identification while maintaining data confidentiality.

**Real-Time Data Analysis:** Rapid identification of diseases requires real-time data processing.

- *Challenge:* Blockchain, particularly public networks, often suffers from latency issues. Ibrahim et al. (Ibrahim et al., 2022) note that transaction speeds in Blockchain need improvement for immediate data analysis in healthcare emergencies.
- *Influence:* Enhancing Blockchain's performance for real-time analysis is crucial for the timely identification and response to new health threats.

**Regulatory and Ethical Compliance:** Disease identification involves managing sensitive health data regulated by laws like

HIPAA and GDPR.

- *Challenge:* Blockchain's immutability conflicts with data privacy regulations that allow data modification or deletion on request. Albassam et al. (Albassam et al., 2024) and Liu et al. (Liu et al., 2020) emphasize the need for frameworks that address these compliance issues within Blockchain networks.
- *Influence:* Ensuring legal and ethical compliance is vital for the practical adoption of Blockchain in healthcare, fostering trust among patients and stakeholders.

**Stakeholder Engagement and Trust:**

Effective identification of unknown diseases requires participation from patients, healthcare providers, and researchers.

- *Challenge:* Vora et al. (2023) highlight the lack of research on building trust and promoting stakeholder engagement in Blockchain systems. Clear education on Blockchain's benefits in disease detection is necessary to encourage data sharing.
- *Influence:* Limited engagement and trust reduce data sharing, impairing the accuracy and effectiveness of disease identification.

**Incentive Mechanisms for Data Sharing:**

Stakeholders are often hesitant to share sensitive health data due to privacy concerns and unclear benefits.

- *Challenge:* Jadhav and Deshmukh (Jadhav & Deshmukh, 2022) indicate that few studies explore incentive models for data sharing in healthcare Blockchain systems. Research into developing rewarding mechanisms (e.g., tokens) is essential to promote data contribution while ensuring privacy.
- *Influence:* Effective incentive mechanisms could broaden participation, improving data quality and diversity for accurate disease identification.

Addressing these challenges will be crucial in developing robust, efficient, and secure Blockchain-based systems for healthcare, ultimately enhancing early detection, disease management, and overall healthcare delivery.

### 3. Material and methodology

#### 3.1. Material

The dataset utilized in this study comprises a structured collection of patient information, focusing on symptoms, medication use, severity, and geographical data. This dataset plays a pivotal role in analysing patterns for early detection of viral diseases.

#### 3.1.1. Overview of the Dataset:

- **Number of Attributes:** The dataset comprises 20 attributes, including various symptoms, medications, severity, date, and area (postal code).
- **Key Features:**

*Symptoms:* The dataset includes key symptoms commonly associated with viral infections, which are crucial for identifying disease patterns. These symptoms encompass: Fever, Fatigue, Dry cough, Difficulty in breathing, Sore throat, Body pain, Nasal congestion, Runny nose, Chills, Diarrhea, Abdominal pain, Anosmia (loss of smell), and Ageusia (loss of taste). Following chart as shown in figure 2 represents the occurrences of different symptoms in patients.

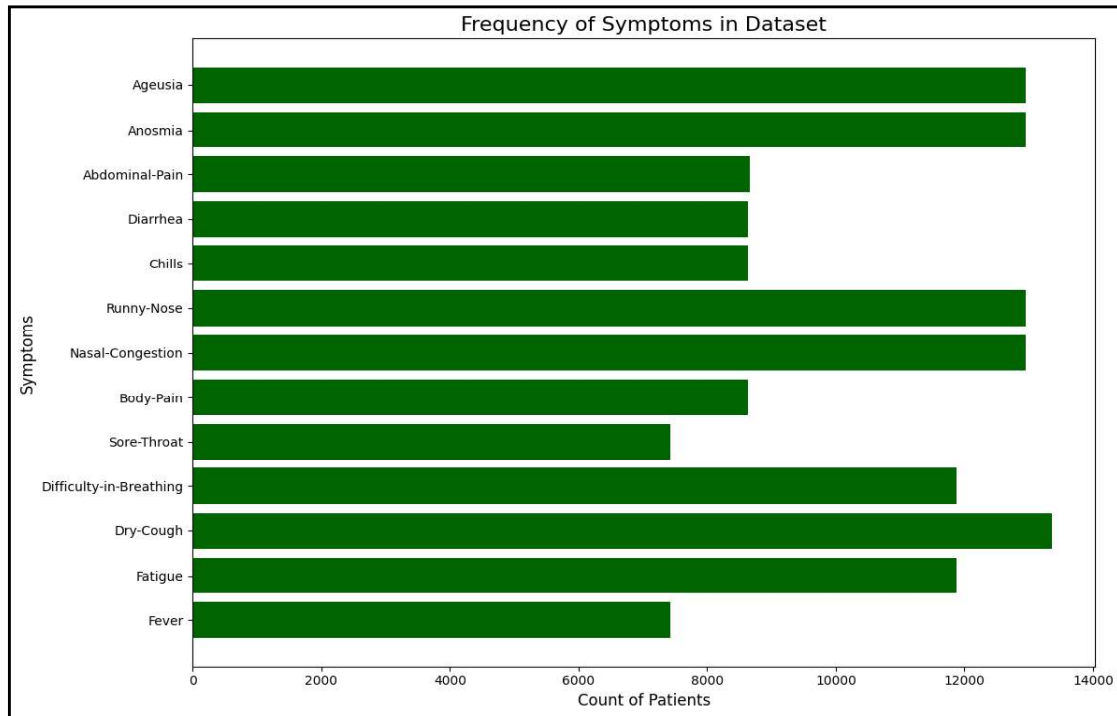
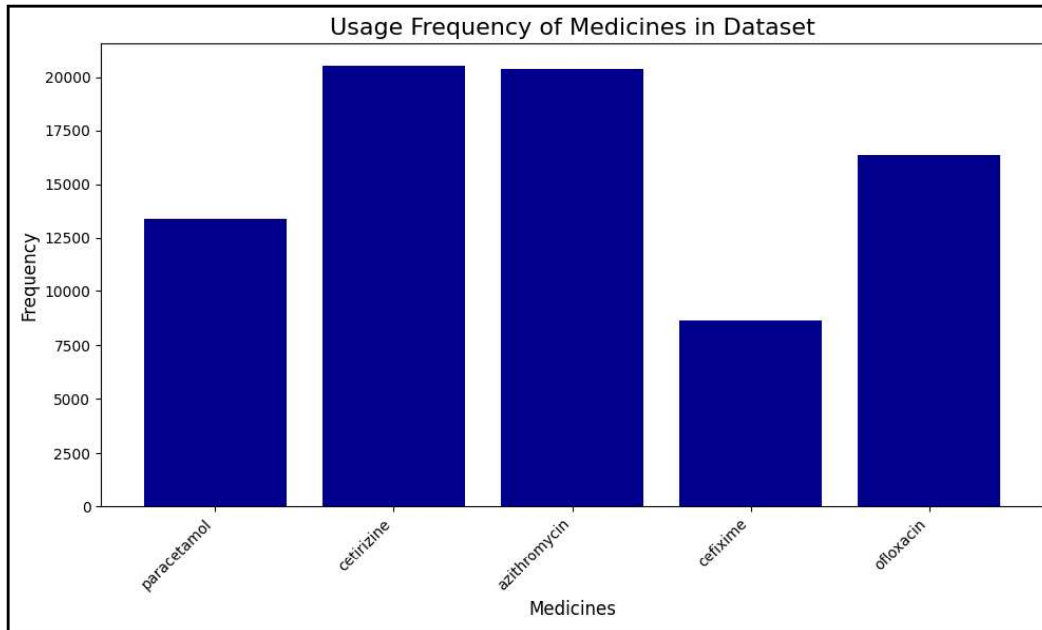


Figure 2. Symptoms occurrences in patients

*Medications:* The dataset also captures details of medications prescribed to patients, including: Paracetamol, Cetirizine, Azithromycin, Cefixime and

Ofloxacin. The bar chart shown in figure 3 illustrates the various medications administered to patients.



**Figure 3.** Medications given to patients

*Date:* Provides a temporal context for each case, aiding in the analysis of symptom progression and outbreak patterns over time.

*Area:* Encodes the geographical location using postal codes, offering insights into regional disease spread and helping identify potential hotspots.

By including a wide range of symptoms, medication data, and both temporal and geographical information, this dataset serves as a robust foundation for conducting in-depth analyses on viral disease patterns. Its structure supports machine learning techniques aimed at enhancing early disease detection and providing informed public health responses.

### 3.1.2. Medicine-Symptom Correlations for Disease Detection

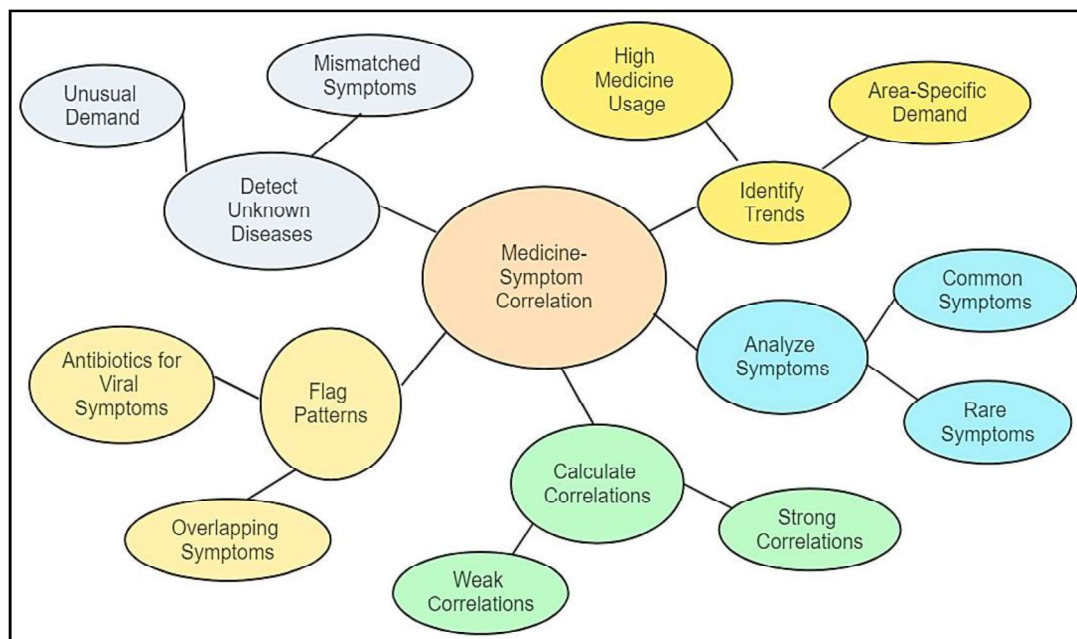
Understanding the relationships between reported symptoms and prescribed medications is crucial for the early detection of diseases. By exploring the correlations between these factors, we can gain valuable insights into developing health patterns and possibly discover new or unusual illnesses. The dataset includes comprehensive information on patient symptoms and the medications they were given, offering a solid foundation to investigate how particular treatments align with various symptom. **Figure 4** illustrates how examining these relationships can aid in spotting emerging health trends and identifying unknown diseases.

Here's an overview of the key processes involved in this analysis:

- **Identifying Trends**
  - *High Medicine Usage* – The system monitors sudden increases in the use of certain medications, such as Paracetamol, Cetirizine, and Azithromycin, which may suggest potential health outbreaks or unusual trends. By linking the rise in medication consumption with specific symptoms like fever, dry cough, or loss of smell and taste, the system can help identify

the early signs of emerging viral conditions.

- *Area-Specific Demand* – By examining regional differences in medication demand through geographic attributes, we can identify specific health concerns or potential outbreaks in localized areas. For example, an increase in Azithromycin prescriptions in a certain region could signal the emergence of a respiratory-related viral outbreak.



**Figure 4.** Understanding dataset

- **Analysing Symptoms**
  - *Common Symptoms* – The most frequently reported symptoms in the dataset, such as fever, dry cough, and difficulty in breathing, are strongly correlated with the use of medications like Paracetamol and Azithromycin. This helps in confirming known viral conditions, especially those linked to respiratory infections, and supports early diagnosis.
  - *Rare Symptoms* – The dataset also monitors less common symptoms like chills and chest discomfort, which can

serve as early indicators of more severe or atypical cases. If these rare symptoms appear along with unusual medication combinations, it could signal the emergence of new viral threats.

- **Calculating Correlations**

*Strong Correlations* – Strong relationship between certain symptoms and medications point to established treatment guidelines. For instance, Paracetamol is commonly used to manage fever, while Azithromycin is often prescribed for respiratory conditions. By tracking how these relationships change over

time, we can continually improve algorithms for detecting diseases.

- *Weak Correlations* – When there are weak or unexpected links between symptoms and medications, it may point to either off-label drug use or the emergence of a new, unidentified illness. For example, if there's a sudden rise in the use of Ofloxacin for treating viral symptoms, this could signal the need for further research into possible secondary bacterial infections or complications that are being addressed with the antibiotic.
- **Flagging Patterns**
  - *Overlapping Symptoms* – When multiple conditions share similar symptoms, like fever in both the flu and COVID-19, the system identifies differences by analysing the full range of symptoms, their intensity, and the treatments involved. This approach enhances the accuracy of differential diagnosis, helping healthcare professionals better differentiate between more common illnesses and rarer ones.
  - *Antibiotics for Viral Symptoms* – The system flags the improper prescription of antibiotics like Cefixime and Ofloxacin for viral symptoms, which may suggest either diagnostic uncertainty or potential co-infections. Identifying these patterns is crucial for promoting accurate treatment and reducing the threat of antimicrobial resistance.
- **Detecting Unknown Diseases**
  - *Unusual Demand* – When there's a noticeable mismatch between medication usage and the usual symptom patterns, the system identifies these irregularities. For instance, if there's a sudden rise in the use of Cetirizine without a corresponding increase in allergy-related symptoms, it could indicate the onset of a viral illness that presents symptoms similar to allergies.

- *Mismatched Symptoms* – If the reported symptoms don't align with the established patterns for a specific medication, the case is marked for further review. This process is essential for uncovering new conditions or variations of known diseases, allowing for quicker action in addressing potential public health risks.

The data provided can be used to identify patterns that help in detecting new or unknown diseases by analysing the trends in symptoms and the usage of medications.

### 3.2. Methodology

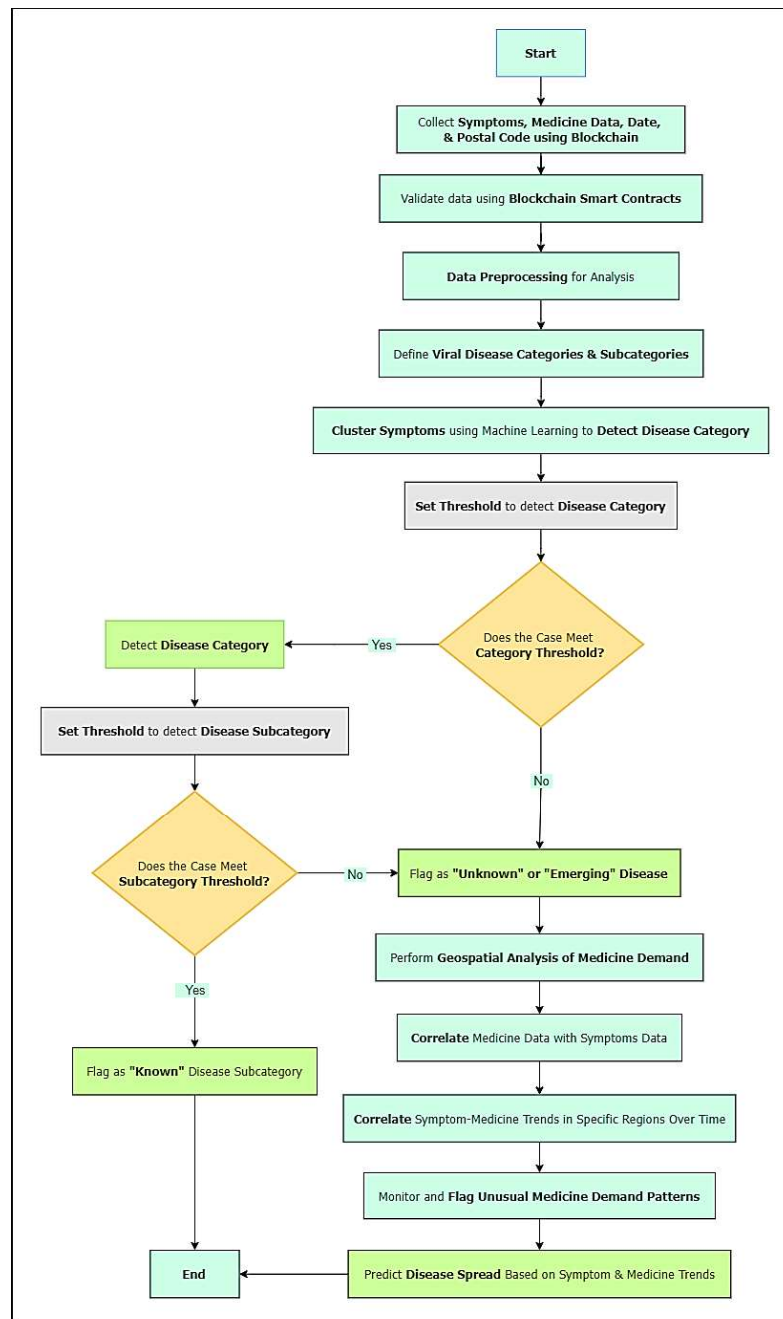
**Figure 5** demonstrates a comprehensive workflow for the identification of unknown viral disease, leveraging Blockchain for data integrity and machine learning for pattern detection. It emphasizes early detection using patient symptoms, medicine demand trends, geographical regions (postal codes), and periodic parameters (dates). The methodology begins by initiating the process of data collection and analysis for early disease detection.

Here's the step-by-step breakdown:

- *Collect Symptoms, Medicine Data, Date, & Postal Code using Blockchain* – At this stage, patient-related data is gathered, which includes various aspects like symptoms—such as fever, cough, and shortness of breath—along with details about the medications prescribed, including types and dosages. The timeline of when these symptoms and treatments were recorded is also noted, alongside geographical data in the form of postal codes to help track regional trends. Blockchain technology is employed to securely store this information in a decentralized, tamper-proof system, ensuring the data remains unchangeable and accessible only to authorized parties.

*Validate Data using Blockchain Smart Contracts* – Blockchain smart contracts are used to automatically validate the collected data, ensuring its accuracy and completeness. They play a crucial role in confirming the integrity of the data by checking that it hasn't been tampered

with. Only after the data passes this validation process is it stored and utilized for further analysis. This step is vital for maintaining the reliability and trustworthiness of the information used in the healthcare system.



**Figure 5.** Methodology

- **Data Pre-processing for Analysis** – The data undergoes pre-processing to ensure it's ready for analysis. This process includes following key steps:
  - *Data cleaning*, which involves removing errors, inconsistencies, or incomplete entries.
  - *Data Transformation*, where the information is standardized to ensure it's in a format that can be effectively used by machine learning algorithms.
  - *Data Reduction*, which simplifies the dataset by eliminating irrelevant details, focusing only on important variables.

This pre-processing ensures the data is of high quality, minimizing the risk of errors in further analysis.

- **Define Viral Disease Categories & Subcategories** – The system groups viral diseases into broader categories, such as respiratory or gastrointestinal viruses. Within these larger categories, it further breaks them down into specific subcategories based on the type of disease, like COVID-19 or influenza. This classification helps the machine learning model detect patterns and accurately identify known diseases based on the symptoms presented.
- **Cluster Symptoms using Machine Learning to Detect Disease Category** – Machine learning models are applied to the dataset to group or cluster similar symptoms. This process helps the system identify common symptom patterns and classify diseases into broader categories based on shared characteristics in patient data. For example, if a group of patients shows symptoms like coughing, fever, and difficulty breathing, the system might categorize the illness

as a "respiratory virus." This approach improves disease classification by recognizing symptom similarities across patients.

- **Set Threshold to Detect Disease Category** – A threshold is set to determine how closely the patient's symptoms match the predefined disease categories. This threshold serves as a confidence level:
  - *If the symptoms meet the threshold, the system assigns the disease to a category.*
  - *If not, it triggers further investigation.*

The threshold ensures that the system is making accurate classifications, reducing the likelihood of misdiagnosis.

- **Decision: Does the Case Meet the Category Threshold?** – This is a decision point:
  - *Yes:* If the symptoms match a known disease category, the process moves forward to the next step, where a specific subcategory is determined.
  - *No:* If the symptoms don't meet the threshold for any known category, the case is flagged as "Unknown" or "Emerging", suggesting it could be a new or unfamiliar disease. The system moves to further investigation.
- **Set Threshold to Detect Disease Subcategory** – Once a disease category is identified, the system establishes an additional threshold to determine the specific subcategory of the disease. This process helps refine the diagnosis, narrowing it down to a particular illness, such as COVID-19 or influenza.

- *Decision: Does the Case Meet the Subcategory Threshold?* – At this decision point:
  - *Yes:* If the symptoms match the subcategory threshold, the disease is classified as a known disease within that category.
  - *No:* If the threshold isn't met, the case is once again flagged as unknown or emerging.
- *Flag as "Unknown" or "Emerging" Disease* – If the system fails to classify the disease under any known category or subcategory, the case is flagged as “Unknown” or “Emerging”. This triggers further analysis and a closer examination of the data to explore the possibility of a new disease.
- *Perform Geospatial Analysis of Medicine Demand* – The system performs a geospatial analysis of medicine demand across different regions using the area/postal code data. This helps:
  - Identify areas with increased demand for specific medications.
  - Detect unusual patterns that could indicate the start of an outbreak in specific geographical regions.
- *Correlate Medicine Data with Symptoms Data* – The system then monitors how these correlations change over time and in various geographical areas. For example, if a region experiences an increase in both specific symptoms and the demand for certain medications, it may signal the onset of a new health concern.
- *Correlate Symptom-Medicine Trends in Specific Regions over Time* – The system tracks how symptoms and medicine demand

evolve over time in specific regions, identifying any patterns that could suggest the spread of a disease. For example, if a region sees both rising symptoms and increased demand for fever medication, it might indicate a growing outbreak.

- *Monitor and Flag Unusual Medicine Demand Patterns* – The system constantly tracks medication demand, keeping an eye out for any unusual patterns. If it detects anomalies, such as sudden increases in drug usage, these are flagged for further investigation, as they could indicate a potential outbreak or unusual health trends.
- *Predict Disease Spread Based on Symptom & Medicine Trends* – Finally, the system uses the gathered data to predict the spread of the disease. By analysing how symptoms and medicine demand change over time and across regions, it forecasts where the disease could spread next.

The methodology concludes by either flagging a disease as a known disease or raising an alert for an unknown or emerging disease. The system's proactive detection capabilities allow healthcare systems to react quickly and effectively to new health threats, potentially containing outbreaks before they become widespread.

#### 4. Layered System Architecture Design

Figure 6 illustrates a multi-layered architecture designed to detect unknown viral diseases in their early stages by integrating Blockchain and Machine Learning. This system is organized into four main layers—Frontend, Backend, Blockchain, and Machine Learning—each playing a unique role in creating a secure, privacy-preserving framework. With a focus

on security, privacy, and seamless real-time data sharing, these interconnected layers work together to enable secure, efficient, and timely data processing and analysis. Here's a detailed breakdown of each layer, outlining their functions, data flows, and the complete path data takes through the system.

- **Frontend Layer**

The frontend layer serves as the main interface where users interact with the system. It's designed to meet the needs of various stakeholders—doctors, patients, pharmacies, distributors, manufacturers, and suppliers—each with custom interfaces. Through this layer, users connect with the backend and Blockchain via APIs and protocols like Web3, JSON RPC, and REST. The frontend layer serves different roles for various stakeholders in the system, ensuring each user has access to the specific tools and information they need. Here's how it works for each group:

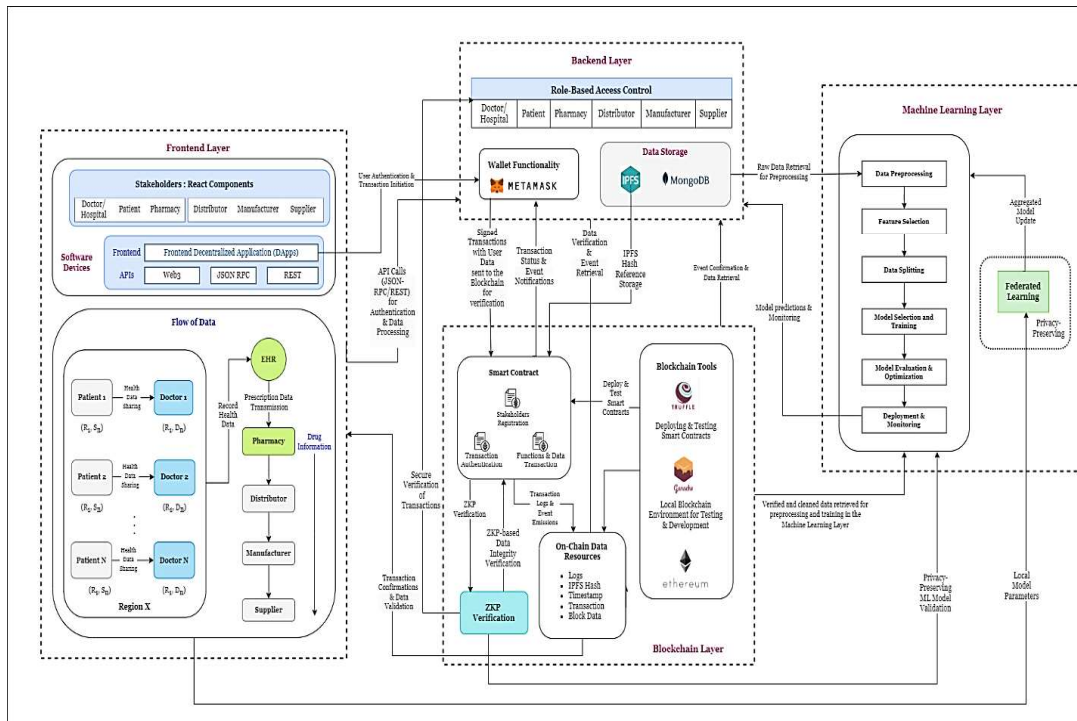
- *Patients*: They use the frontend to securely submit their health data, such as symptoms and test results, to their doctors.
- *Doctors*: The frontend allows doctors to access and review patient data, update medical records, and add important health information to the system.
- *Pharmacy*: Pharmacies can view and process prescription data, ensuring they dispense the right medications to patients. The pharmacy transmits data concerning medications to the supply chain.
- *Distributors, Manufacturers, and Suppliers*: In addition to having their own medicine inventory, these stakeholders are able to visualize the supply chain's medication demand.

Each group interacts with the system through tailored interfaces, ensuring that the frontend is customized to meet their unique needs while supporting seamless collaboration and data sharing.

**Data Flow:**

- The data flow begins when a patient in region  $R_N$ , experiencing symptoms  $S_N$ , visits doctor  $D_N$ . ( $R_N, S_N, \rightarrow, D_N$ ). During the visit, the patient shares their health information, such as symptoms, with the doctor. In a broader context, patients within the same region (*Region X*) share their health data with their respective doctors, contributing to a pool of real-time information that lays the groundwork for early disease detection.
- *Doctors and EHR (Electronic Health Record)*: Once doctors receive data from patients, they update the EHR system with this information. The EHR serves as a central hub for storing and organizing health data, providing a structured system for easy access to patient information.
- *Prescription Data Transmission*: Following patient assessments, doctors prescribe medications or treatments. These prescriptions are then sent through the EHR system to pharmacies and other relevant stakeholders, ensuring smooth communication and accurate processing.
- *Pharmacy*: Pharmacies receive prescription data from the EHR system, which allows them to access patient-specific prescription details. This ensures that pharmacies dispense the correct medications based on the doctor's authorized prescriptions.
- *Drug Information*: Pharmacies also have access to up-to-date drug information, which flows downstream to other stakeholders involved in drug production and distribution. This ensures everyone in the supply chain stays informed about the latest drug developments.
- *Supply Chain Stakeholders*: From the pharmacy, data continues to flow downstream to various participants in the supply chain:

- **Distributor:** Manages the distribution of medications to pharmacies and healthcare facilities, ensuring timely delivery.
- **Manufacturer:** Oversees the production of drugs, ensuring they are available as needed and meet regulatory standards.
- **Supplier:** The final link in the supply chain, the supplier ensures that raw materials are provided to manufacturers according to demand.



**Figure 6.** System Architecture Diagram

- **Backend Layer**

The backend layer acts as the control centre of the system, overseeing data management and ensuring secure access. It enforces Role-Based Access Control (RBAC), handles data storage, and authenticates all transactions to maintain security and privacy. Here's how it functions:

- **Role-Based Access Control (RBAC):** Role-Based Access Control (RBAC) governs data access by assigning permissions according to the user's specific role. For instance, only medical professionals such as doctors have the authority to view patient records, while pharmacies are

granted access to prescription details. This structure ensures that sensitive information is restricted to authorized individuals, thereby bolstering both privacy and security.

- **Wallet Functionality:** MetaMask is used for user authentication and signing transactions. Each user has a digital wallet that securely interacts with the Blockchain layer, ensuring transparent and trustworthy management of data.
- **Data Storage:** Data is stored in MongoDB for quick access and in IPFS (InterPlanetary File System) for secure, immutable references.

MongoDB provides efficient data retrieval, while IPFS guarantees that each piece of data is stored with unique hash references, ensuring integrity and long-term reliability.

The backend processes API requests from the frontend, handling authentication and data transactions. After validation, data and transaction references are securely transferred to the Blockchain for further handling and storage, maintaining security throughout the system.

- **Blockchain Layer**

The Blockchain layer is the backbone of the system's data security and transparency. It uses Smart Contracts and Zero-Knowledge Proofs (ZKP) to maintain data integrity while preserving user privacy.

- *Smart Contracts*: Smart contracts are pre-defined codes on the Blockchain that govern stakeholder actions, such as registering users, authenticating transactions, and handling data sharing. They execute autonomously to ensure that transactions happen securely and reliably.
- *Zero-Knowledge Proof (ZKP) Verification*: Zero-Knowledge Proof (ZKP) is a cryptographic technique that enables the validation of a statement's truthfulness without disclosing any extra details beyond the statement's accuracy. This method ensures that data can be verified while keeping the underlying information confidential. In healthcare systems, ZKP is particularly valuable as it ensures data integrity while safeguarding sensitive medical information, thereby supporting privacy and security in data verification processes.
- *Blockchain Tools*: Blockchain tools are essential for developing, testing, and securely deploying smart

contracts. Truffle serves as a development framework, allowing developers to write, test, and deploy smart contracts efficiently, while Ganache provides a local Ethereum Blockchain environment for testing in a controlled setting without transaction fees. Together, Truffle and Ganache enable rapid iteration and debugging of contracts before live deployment. The Ethereum network then acts as the main Blockchain platform, supporting the transparency and security of the system. Using Ethereum's Blockchain explorer, stakeholders can view and verify transactions, ensuring trust and accountability. This integration of tools creates a secure, transparent system for handling health data in a decentralized manner.

- *Data Storage and Retrieval*: Verified data transactions are stored on-chain, which includes transaction logs, IPFS hashes, timestamps, and other reference data. This data trail is essential for transparency, as it provides an auditable log of every action.
- *Data Flow*: When a transaction (e.g., patient health update) is initiated from the frontend, it's processed in the Blockchain layer. The transaction undergoes ZKP-based verification, and event logs are stored on-chain, creating a trustworthy data environment.

- **Machine Learning Layer**

The machine learning layer processes health data to identify early signs of viral diseases, using a series of stages to clean, analyse, and validate the data—all while safeguarding user privacy.

- *Data Processing Steps*:
  - *Data Pre-processing*: Raw health data is cleaned and standardized, ensuring

consistency and eliminating irrelevant information.

- *Feature Selection*: Important health metrics are identified, allowing the model to focus on the most relevant data points that can help detect disease patterns.
- *Data Splitting*: The data is divided into training and testing sets, enabling the model to learn from one portion while validating its accuracy on the other.
- *Model Training and Optimization*: The system explores various machine learning algorithms to uncover patterns that might indicate the presence of an unknown viral disease. The model is carefully fine-tuned for precision and reliability. Once trained, it's deployed to monitor real-time health data, alerting stakeholders when early signs of a disease are detected.
- *Federated Learning*: Federated learning enables the model to train on data stored locally on users' devices, keeping raw data private. The main purpose of connecting the Frontend Layer to the Federated Learning system is to enable health data collection from patients in a way that preserves privacy and allows for decentralized machine learning.

#### **Data Flow:**

- Once trained, the model analyses new data, sending predictions to the backend. The Blockchain layer also feeds verified, up-to-date data into the model, keeping it aligned with current health trends.

This flow of data creates a smooth and interconnected path from patient health information to healthcare providers, and extends all the way through the supply chain. By involving each stakeholder in a

decentralized manner, the system guarantees secure data handling, with real-time updates flowing throughout the healthcare ecosystem. This setup boosts early detection of potential health risks, especially by aggregating data from multiple patients in a region and monitoring medication usage through pharmacies and the supply chain.

## **5. Model Development**

### **5.1. Identifying Symptoms by Category and Subcategory**

Viral diseases are systematically grouped into categories and subcategories (as shown in figure1) based on the organ systems they affect and the nature of their symptoms. For example, Respiratory Viral Infections include subcategories like the Common Cold, Influenza, and COVID-19, all of which primarily impact the respiratory tract. Symptoms such as cough, nasal congestion, and breath-ing difficulties are common in this category. Similarly, Exanthematous Viral Infections, character-ized by skin rashes, include diseases like Measles, Chickenpox, and Rubella. Other categories in-clude Gastrointestinal Viral Infections—comprising subcategories like Norovirus and Rotavirus—which are associated with symptoms such as diarrhea and vomiting. Systemic Viral Infections, such as Dengue Fever and Chikungunya, often present with generalized symptoms like body aches and fever. Localized infections, such as Genital Viral Infections and Oral Viral Infections, include con-ditions like Herpes Simplex Virus (HSV), which cause localized sores and discomfort. Arboviral Infections, transmitted by arthropods, cover diseases like Zika Virus and Yellow Fever, while Hem-orrhagic Viral Fevers, such as Ebola and Marburg Virus, are known for severe systemic symptoms, including bleeding disorders.

Each category's specific symptoms help differentiate and diagnose these diseases. For example, in the Respiratory Viral Infections

category, a mild fever and runny nose may suggest the common cold, while severe muscle aches and high fever may indicate influenza. Similarly, rashes starting on the face in Measles differ from the itchy blisters of Chickenpox.

## 5.2. Symptom Weight Assignment

Each symptom is weighted based on its relevance to specific diseases, factoring in prevalence, severity, and clinical importance. For instance, fever is highly relevant to influenza, so it might have a weight of 0.35 in the scoring system. The total weight for all symptoms in a disease

$$\text{Normalized Weight} = \frac{\text{Raw Weight}}{\text{Sum of All Raw Weights in the Category or Subcategory}}$$

For example,

Suppose raw weights for Respiratory Diseases:

Dry Cough = 0.3, Difficulty in Breathing = 0.5, Fever = 0.2.

Total Raw Weight = 0.3 + 0.5 + 0.2 = 1.0 (Sums to 1).

### 5.3.2. Category and Subcategory Score Calculation

Category scores aggregate the symptom scores for each group. The system identifies the category with the highest alignment to the user's symptoms. Once the best-matching category is determined, subcategories within that category are scored similarly to pinpoint the most relevant sub-condition. For example, if the Respiratory Viral Infections category has a score of 0.428, exceeding the threshold of 0.3, it is flagged as relevant. Within this category, subcategories like influenza and the common cold are scored to identify the best match.

Given symptoms: *Dry Cough (1)*, *Difficulty Breathing (0)*, and *Fever (1)*.

Normalized weights for Respiratory Diseases:

category sums to 1, ensuring balanced contributions in the diagnostic process.

## 5.3. Scoring Systems for Diagnosis

### 5.3.1. Symptom Score Calculation

Scores are calculated by summing the weights of symptoms selected by the user. For instance, if "dry cough" (weight 0.3), "fever" (0.2), and "difficulty breathing" (0.5) are reported, the total score is 1.0. This score is normalized and compared against a threshold to determine if the symptoms align strongly with a specific disease.

$$\begin{aligned} \text{Dry Cough} &= 0.238, \text{ Difficulty Breathing} \\ &= 0.333, \text{ Fever} = 0.190. \\ \text{Score} &= \sum (\text{Weight of Symptom} \times \\ &\text{Presence Indicator}) \dots (1) \\ \text{Score} &= (0.238 \times 1) + (0.333 \times 0) \\ &\quad + (0.190 \times 1) = 0.428 \end{aligned}$$

### 5.4. Thresholds for Category / Subcategory Classification

Threshold score calculation serves as a benchmark to determine the likelihood of diagnosing a disease based on symptom matching. A percentage threshold is used to filter diseases, evaluating how closely the user's symptoms align with predefined profiles. Once the best-matching category is identified by meeting or exceeding the threshold, subcategories within that category are further assessed. If the score surpasses 60-70%, it indicates a strong match, suggesting the category and its associated

disease are likely. Similarly, subcategories are evaluated using their respective thresholds to refine the diagnosis. However, if a score falls below 40%, the match is considered weak, making it difficult to pinpoint a specific disease. In such cases, the disease is classified as "unknown" but may still belong to a category with the most aligned symptoms. Additional diagnostic

tests or more detailed symptom inputs would then be necessary to reach a clearer conclusion.

Once the best-matched category is identified, the next step is to find the subcategory within that category that best matches the user's symptoms.

$$\text{Best Matched Category} = \max(\text{Category}_{\text{Score}}) \quad \dots(2)$$

$$\text{Best Matched Subcategory} = \max(\text{SubCategory}_{\text{Score}}) \quad \dots(3)$$

$$\text{If } [(Category_{\text{Score}} > Category_{\text{Threshold}}) \rightarrow Category_{\text{Relevant}}] \quad \dots(4)$$

$$\text{If } [(Subcategory_{\text{Score}} > Subcategory_{\text{Threshold}}) \rightarrow Subcategory_{\text{Relevant}}] \quad \dots(5)$$

Establish a threshold score for each disease category (e.g., Respiratory Diseases). If the calculated score for a category surpasses this threshold, it is flagged as relevant for further assessment. For example:

Threshold for Respiratory Diseases = 0.3  
 Calculated score for Respiratory Diseases = 0.428.  
 Since  $0.428 > 0.3$

When symptoms don't strongly match any predefined conditions, the system flags the possibility of an unknown disease. By analysing symptom-drug correlations, it identifies potential treatment gaps. If symptoms like "fatigue" or "dry cough" are unmapped to any drugs, this suggests the presence of a new or rare condition.

## 5.5. Model Performance

### 5.5.1. Category Model Evaluation

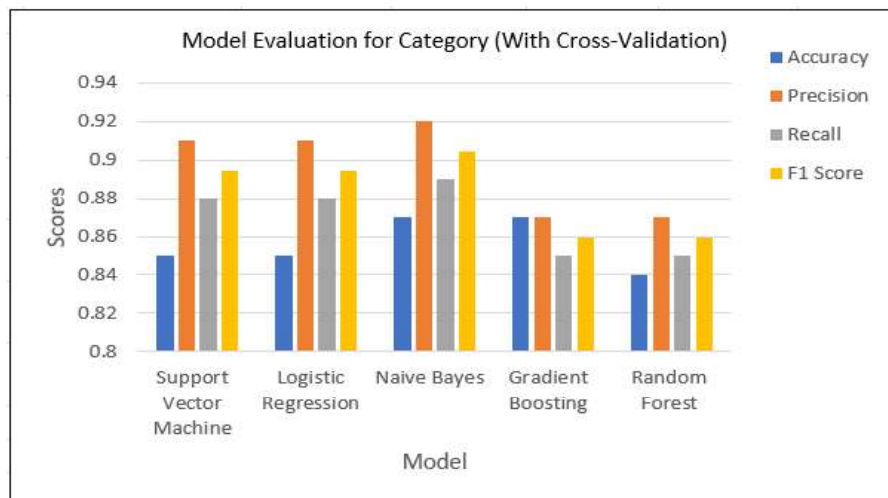
The category Respiratory Diseases is marked as relevant if its score exceeds the threshold. If the score is lower, it is considered irrelevant, indicating that the user provided fewer symptoms, making it harder to predict a specific disease (equation 4).

The performance of the models used for category prediction was assessed through a variety of metrics: *accuracy*, *precision*, *recall*, and *F1-score* (as shown in table 5), all after applying cross-validation. Logistic Regression achieved a mean cross-validation accuracy of 0.85, with precision and recall scores of 0.91 and 0.88, respectively. This indicated a slight decline in performance compared to its initial results, as shown in Table 8.2. On the other hand, Naive Bayes demonstrated better consistency, reaching a mean cross-validation accuracy of 0.87, along with steady precision (0.92) and recall (0.89). This made it one of the most reliable models. Gradient Boosting also displayed stable performance, with a mean cross-validation accuracy of 0.87 and well-balanced precision and recall values of 0.87 and 0.85, respectively.

After a category is identified as relevant, the next step is to evaluate the subcategories within that category. Each subcategory (such as Flu, Common Cold, or COVID-19) is assigned a specific threshold score. If the calculated score for a subcategory surpasses its threshold, that subcategory is marked as relevant within the parent category (equation 5). The subcategory is selected based on the best-matched category. If no suitable subcategory is found, it will be classified as "unknown". The subcategory is always searched within the top-matching category, as outlined in equations (2) and (3).

**Table 5.** Model Evaluation for Category

| Model Name             | Accuracy | Precision | Recall | F1 Score |
|------------------------|----------|-----------|--------|----------|
| Support Vector Machine | 0.85     | 0.91      | 0.88   | 0.894    |
| Logistic Regression    | 0.85     | 0.91      | 0.88   | 0.894    |
| Naive Bayes            | 0.87     | 0.92      | 0.89   | 0.904    |
| Gradient Boosting      | 0.87     | 0.87      | 0.85   | 0.859    |
| Random Forest          | 0.84     | 0.87      | 0.85   | 0.859    |



**Figure 7.** Model Evaluation for Category

While Random Forest had a slightly lower mean cross-validation accuracy of 0.84, it still exhibited solid performance, with precision and recall scores of 0.87 and 0.85. These results highlighted Multinomial Naive Bayes and Gradient Boosting as the top contenders for generalizing to new, unseen data, while Random Forest showed strong robustness across all metrics, as illustrated in the bar graph in Figure 7.

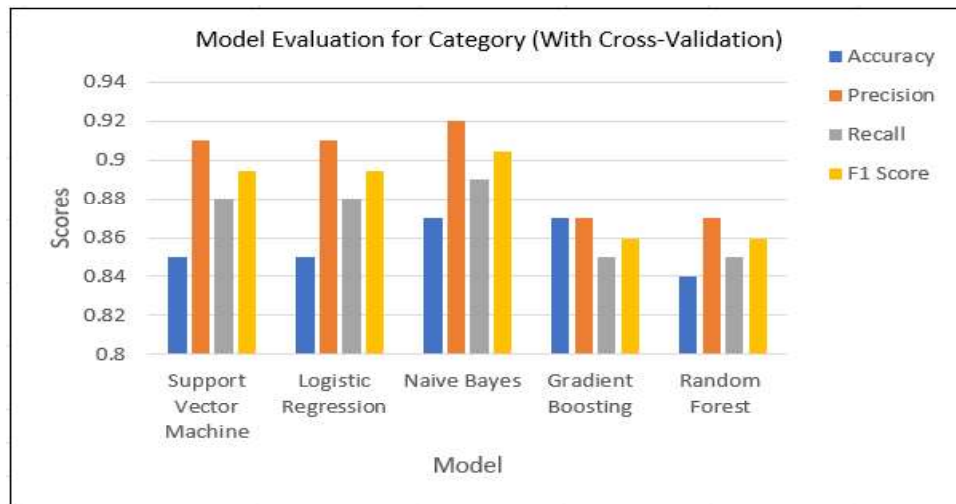
### 5.5.2. Subcategory Model Evaluation

Figure 8 provides a graphical representation

of the results. The evaluation of subcategory prediction models, following cross-validation, revealed varying levels of performance, showcasing their ability to generalize on new, unseen data, as seen in Table 6. Both SVM and Logistic Regression achieved identical mean cross-validation accuracies of 0.65, with precision scores of 0.70 and recall values of 0.69. These results indicate a stable and reasonable performance for subcategory prediction, with both models performing similarly. Figure 8 highlights graphical representation of the results.

**Table 6.** Model Evaluation for Subcategory

| Model Name             | Accuracy | Precision | Recall | F1 Score |
|------------------------|----------|-----------|--------|----------|
| Support Vector Machine | 0.65     | 0.7       | 0.69   | 0.694    |
| Logistic Regression    | 0.65     | 0.7       | 0.69   | 0.694    |
| Naive Bayes            | 0.67     | 0.72      | 0.72   | 0.72     |
| Gradient Boosting      | 0.63     | 0.66      | 0.67   | 0.664    |
| Random Forest          | 0.62     | 0.62      | 0.64   | 0.629    |



**Figure 8.** Model Evaluation for Subcategory

However, Multinomial Naive Bayes emerged as the top performer, delivering the highest mean cross-validation accuracy of 0.67, along with well-balanced precision and recall scores of 0.72. This model demonstrated the best reliability for subcategory classification. On the other hand, Random Forest showed the lowest accuracy at 0.62, with a precision of 0.62 and recall of 0.64, suggesting it struggled to generalize effectively.

In summary, these findings position Naive Bayes as the most dependable model for subcategory prediction after cross-validation, while the other models displayed varying

levels of overfitting or underperformance.

## 6. Unknown Diseases Detection through Symptom Patterns and Drug Utilization

This study proposes a novel approach to detect potential unknown diseases by analysing patterns of symptoms, feature engineering, and drug utilization and demand. The method combines traditional symptom classification with modern machine learning techniques and drug demand patterns, focusing on the correlations between symptoms, prescribed

medicines, and the possibility of identifying emerging or unknown diseases.

### 6.1. Symptom and Feature Engineering

The symptoms in the dataset were categorized into five primary feature groups based on common clinical knowledge and prior research:

- *Nasal Symptoms*: Includes symptoms related to nasal congestion and runny nose.
- *Respiratory Symptoms*: Encompasses dry cough, difficulty in breathing, sore throat, and fatigue.
- *Sensory Loss*: Includes anosmia (loss of smell) and ageusia (loss of taste).
- *Flu Index*: Captures general flu-like symptoms such as fever, fatigue, chills, and body pain.
- *Gastrointestinal Symptoms*: Focuses on gastrointestinal distress, including diarrhea and abdominal pain.

This categorization groups similar symptoms into broader, clinically meaningful classes, allowing for easier analysis and pattern recognition.

### 6.2. Medicine Utilization and Demand

The study also takes into account drug utilization based on reported prescriptions, focusing on five key medications:

- *Paracetamol*: Often prescribed for pain and fever relief.
- *Cetirizine*: Commonly used for treating allergic reactions and nasal symptoms.
- *Azithromycin*: An antibiotic frequently prescribed for respiratory infections.
- *Cefixime*: Another antibiotic used for treating bacterial infections.

- *Ofloxacin*: Used for treating respiratory and urinary tract infections.

By correlating these medications with the symptoms in the dataset, common treatments for each symptom cluster are identified, helping to hypothesize the likely underlying disease.

## 7. Conclusion and Future work

This research presents a novel Blockchain-based framework integrated with Machine Learning to address critical gaps in healthcare, specifically in early disease detection and supply chain management. A comprehensive literature survey revealed the synergistic potential of Blockchain for secure, decentralized data management and Machine Learning for predictive analytics in healthcare. Building on these insights, a multi-layered architecture was developed to incorporate decentralized symptom reporting, secure patient data storage, and real-time analysis of medicine demand trends. The implemented system demonstrated robust performance in identifying disease categories and subcategories, achieving 87% accuracy for disease categories and 67% for subcategories using models like Multinomial Naive Bayes and Gradient Boosting. Additionally, the framework effectively flagged unknown diseases through symptom-medication correlation anomalies, ensuring its applicability in managing emerging health threats. The integration of geospatial and temporal analyses enabled precise tracking of outbreaks and resource allocation. Blockchain's immutability and transparency ensured secure data sharing while fostering trust among stakeholders. Overall, this work establishes a scalable and adaptive system that enhances healthcare resilience by combining predictive analytics with transparent supply chain management. The results demonstrate the feasibility of transforming healthcare systems to efficiently manage viral disease outbreaks.

and safeguard public health.

Despite its promising outcomes, the framework has limitations that require attention. Scalability remains a challenge due to the storage constraints of Blockchain, which can be addressed by integrating off-chain solutions like IPFS. While the system ensures data privacy, further improvements such as federated learning and advanced cryptographic techniques (e.g., zero-knowledge proofs) are essential for privacy-preserving analytics. Interoperability with existing healthcare systems is another area for improvement, requiring the adoption of standardized data exchange protocols such as HL7 FHIR. Real-time data integration from

IoT devices, including wearable health monitors, could enhance symptom tracking and outbreak prediction. Additionally, incentive mechanisms, such as token-based systems, could encourage broader stakeholder participation in data sharing. Future work will also focus on refining Machine Learning models for better subcategory prediction accuracy and expanding the framework to support the identification of rare and complex diseases. These enhancements aim to create a robust, real-time healthcare system capable of mitigating future pandemics and ensuring equitable healthcare access globally.

## References:

- Albassam, N., Xu, S., & Liu, X. (2024). Integration of blockchain and cloud computing in telemedicine. *PLOS ONE*, *19*(1), Article e0296644.
- Ali, M. S., Vecchio, M., & Pincheira, M. (2021). A survey on blockchain-based self-sovereign patient identity in healthcare. *IEEE Access*, *9*, 90478–90494.
- Banerjee, R. B., et al. (2020). Measles: Clinical manifestations and diagnosis. *The Journal of Pediatrics*, *205*, 13–22.
- Centers for Disease Control and Prevention. (2023). *The common cold: Information for healthcare professionals*. <https://www.cdc.gov/common-cold/hcp>
- Farley, S. S., & [ostali autori]. (2020). Genital herpes: Diagnosis and management. *Journal of Clinical Microbiology*, *58*(2), 251–257.
- Feldmann, H., & Geisbert, T. W. (2011). Ebola haemorrhagic fever. *The Lancet*, *377*(9768), 849–862.
- Glass, A. R., & [ostali autori]. (2017). Rotavirus infections in children: Clinical management and preventive strategies. *Pediatrics*, *140*(4), Article e20171823.
- Gupta, S., & Malhotra, V. (2021). Blockchain for IoT in healthcare data management. *Computers in Biology and Medicine*, *120*, Article 103715.
- Gupta, S., Malhotra, V., & Singh, S. N. (2023). Securing IoT-driven remote healthcare data through blockchain. In *Advances in Data and Information Sciences* (pp. 47–56). Springer.
- Houtan, M., Hafid, A., & Makrakis, D. (2020). A survey on blockchain-based self-sovereign patient identity in healthcare. *IEEE Access*, *8*, 90478–90494.
- Ibrahim, A. S., Ali, M. A., & Mahmoud, T. S. (2022). Machine learning algorithms for medicine demand forecasting. *IEEE Transactions on Biomedical Engineering*, *69*(3), 457–467.
- Jadhav, J. S., & Deshmukh, J. (2022). A review study of the blockchain-based healthcare supply chain. *Social Sciences & Humanities Open*, *6*(1), Article 100328.

- Jadhav, J. S., & Deshmukh, J. (2024). Synergizing machine learning and blockchain for pioneering early disease detection: A focused study on COVID-19 diagnosis. *Journal of Medical Diagnostic Methods*, 13(3), Article 481.
- Jadhav, J. S., & Deshmukh, J. (2025). Advancing machine learning in COVID-19 diagnostics: Symptom-based classification and ensemble techniques. *South Eastern European Journal of Public Health*, 3044–3061
- Johnson, D. C. (2023). Chickenpox: A review of disease management. *BMJ*, 358, 482–489.
- Kumar, R., Kumari, S., & Gupta, M. K. (2024). Blockchain technology in healthcare: A systematic review. *Healthcare*, 12(2), 1–16.
- Liu, Y., Li, L., & Wang, Z. (2022). Smart contracts for automated healthcare supply chain transactions. *IEEE Access*, 11, 12345–12356.
- Liu, Z., Zhang, H., & Wu, J. (2020). A user revocation scheme in CP-ABE systems for healthcare applications. *Journal of Medical Systems*, 44(1), Article 14.
- Naruse, T., Saito, Y., & Akiyama, M. (2022). Attribute-based encryption with attribute revocation and grant function. *IEEE Transactions on Information Forensics and Security*, 17(4), 1234–1247.
- Radanović, P., & Likić, R. (2018). Opportunities for use of blockchain technology in medicine. *Applied Health Economics and Health Policy*, 16(5), 583–590.
- Salah, F., Vecchio, J., & Fahmy, H. (2019). Blockchain for AI: Review and open research challenges. *IEEE Access*, 7, 10127–10149. <https://doi.org/10.1109/ACCESS.2018.2890507>
- Vora, A., Jabbar, A., & Chamola, A. (2023). Blockchain for healthcare supply chain management: A review. *Frontiers in Public Health*, 11, Article 897321.
- Wen, J., Wang, L., & Han, F. (2020). Blockchain technology in healthcare data management: Security and privacy issues. *IEEE Access*, 8, 15967–15984.
- World Health Organization. (2023). *Coronavirus disease (COVID-19): Symptoms*. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/coronavirus-disease-covid-19-symptoms>
- Zhang, P., Schmidt, D. C., & White, J. (2021). Blockchain technology use cases in healthcare. *Advances in Computing*, 111(4), 1–20.

---

**Jayendra S. Jadhav**

Department of Computer  
Engineering, Rajiv Gandhi  
Institute of Technology,  
University of Mumbai, Mumbai,  
India

[jayendra071985@gmail.com](mailto:jayendra071985@gmail.com)  
ORCID 0000-0001-6767-6580

**Jyoti Deshmukh**

Department of Computer  
Engineering, Rajiv Gandhi  
Institute of Technology,  
University of Mumbai, Mumbai,  
India

[Jyoti.Deshmukh@mctrgit.ac.in](mailto:Jyoti.Deshmukh@mctrgit.ac.in)  
ORCID 0000-0002-6671-3041

---

