

## Data Architecture to Facilitate the Diagnosis of Arboviruses

Vagner Silva, Kauã Lima, Matheus Nascimento, Gabriel Souza, Jean Turet  
Federal University of Alagoas

Group of Engineering in Decision-Making and Artificial Intelligence

### INTRODUCTION & AIM

Arboviruses, particularly dengue, chikungunya and zika, represent significant public health challenges in Tropical and subtropical regions worldwide. Brazil is notably affected, with dengue presenting a persistent threat, with over 3 million cases reported in 2024 alone According to Confederação Nacional de Municípios (CNM, 2024). This highlights the urgency of addressing this issue.

The complexities of arboviruses extend beyond mere transmission; their similar symptoms often lead to diagnostic challenges. Fever, rash and joint pain are common across these diseases, making it difficult for healthcare providers to accurately diagnose and differentiate between them without comprehensive diagnostic tools. This often results in under-reporting, misdiagnosis, and ineffective treatment, exacerbating the burden on the healthcare system and affected populations.

This study proposes a novel system that integrates and standardizes data collected at health posts into a comprehensive data architecture. This integrated system leverages artificial intelligence (AI) algorithms to enhance diagnostic accuracy and help doctors suggest effective treatments for arboviruses.

### METHOD

The methodology of this project was designed to integrate data related to arboviruses such as dengue, chikungunya and zika. Through the development and implementation of standardized data collection protocols to ensure consistency. This includes detailed patient data such as temperature, symptoms, travel history, and other relevant health information. The primary goal was to ensure the quality and efficiency in data processing.

#### 1. Data Collection

Utilize electronic medical records (EMRs) to collect and store patient data, ensuring clinical information is accurately recorded, secure, and easily accessible.

#### 2. Data Storage

Data were stored using hybrid databases systems and object storage such Amazon S3 and Amazon RDS to manage both structured and unstructured data formats. This ensures flexibility in storage and retrieval. (Amazon, 2023)

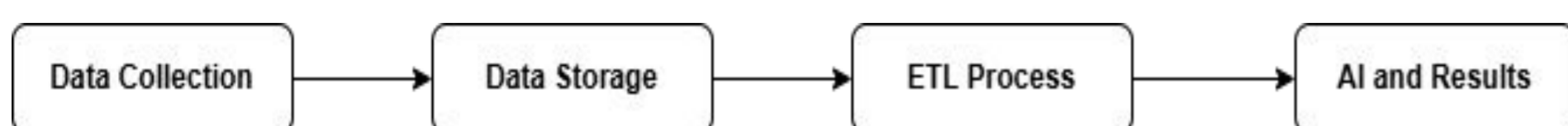
#### 3. ETL Process

Extraction, transformation, and loading process was employed using AWS Glue to develop an automated pipeline to extract data from different sources, transform into a consistent format, and load it into a centralized database. (Amazon, 2023)

#### 4. Machine Learning

Machine learning algorithms were trained to using the collected data to identify patterns and predict arbovirus diagnosis.

The flowchart can explain the process more clearly



### RESULTS & DISCUSSION

#### Impact on Sustainability

By embracing digital solutions like electronic medical records (EMRs), healthcare providers can streamline workflows, enhance data accessibility, and promote ecological sustainability. This transition not only reduces reliance on paper, thereby conserving natural resources, but also establishes a robust, organized framework for managing sensitive health information.

#### Improved Diagnosis

The use of comprehensive datasets collected from health posts, the system integrates multiple diagnostic criteria to achieve a nuanced understanding of symptoms and patterns. This approach minimizes errors, facilitates early intervention, and increases confidence in diagnostic outcome.

#### Improved Decision Making

Harnessing data-driven methodologies enhances the public health systems to respond to outbreaks. By integrating real-time analytics and predictive modeling, decision-makers can allocate resources strategically and prioritize high risk areas, ensuring a more unified and impactful effort against disease spread.

#### Effective Treatment

Incorporating AI-generated insights allows the development of personalized treatment plans, ensuring interventions are precisely tailored to individual patient needs. This approach leverages aggregated clinical data to identify optimal therapeutic strategies, enhancing recovery rates and patient satisfaction.

Advanced diagnostic tools that integrate comprehensive data accuracy and facilitate early interventions. Data-driven approaches empower better decision making in public health, optimizing resource allocation. Additionally, AI-generated information enables a new perspective on treatment plans.

### CONCLUSION

The system highlights the importance of having a robust data collection and management system. By tackling diagnostic complexities and fostering evidence-based treatment protocols, this innovative approach significantly enhances healthcare efficiency and patient outcomes. In addition to its immediate benefits, the system promotes sustainability through paperless processes, conserving natural resources and ensuring secure, organized data handling. By combining technological innovation with public health strategies, this project offers a robust framework for reducing the burden of arboviruses and ultimately saving lives.

### FUTURE WORK / REFERENCES

- Amazon (2023). what is ETL?. Retrieved from <https://aws.amazon.com/pt/what-is/etl/>
- Amazon. (2023). Amazon S3. Retrieved from <https://aws.amazon.com/s3>
- Amazon (2023). Amazon RDS. Retrieved from <https://aws.amazon.com/pt/rds/>
- CNM (2024). Confederação Nacional dos Municípios. Retrieved from <https://cnm.org.br/comunicacao/noticias/dengue-brasil-ja-registra-mais-de-3-milhoes-de-casos-cnm-orienta-gestores>