Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Toward a Unified Smart Alert Framework for Wildlife Disease Surveillance in Brazil through Citizen Science

Jaqueline S. Angelo
Fundação Oswaldo Cruz
Livia S. Abdalla
Fundação Oswaldo Cruz
Douglas A. Augusto
Fundação Oswaldo Cruz
Luisa K. Spolavori
Fundação Oswaldo Cruz
Gerlania F. Rodrigues
Fundação Oswaldo Cruz
Alessandro P. M. Romano
Ministério da Saúde
Daniel G. Ramos
Ministério da Saúde
Pedro Henrique de O. Passos
Ministério da Saúde
Marcia Chame
Fundação Oswaldo Cruz
Eduardo Krempser
eduardo.krempser@fiocruz.br

Fundação Oswaldo Cruz

Research Article

Keywords: Early warning system, disease surveillance, spatiotemporal clustering, index-based, anomaly detection

Posted Date: November 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11694879/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Toward a Unified Smart Alert Framework for Wildlife Disease Surveillance in Brazil through Citizen Science

Jaqueline S. Angelo[1], Livia S. Abdalla[1], Douglas A. Augusto[1], Luisa K. Spolavori[1], Gerlania F. Rodrigues[1], Alessandro P. M. Romano[1,2], Daniel G. Ramos[3], Pedro Henrique de O. Passos[3], Marcia Chame[1], Eduardo Krempser[1*]

[1*]Oswaldo Cruz Foundation, Rio de Janeiro, RJ, Brazil.
[2]General Coordination of the Strategic Health Surveillance Information Center of the Ministry of Health, Brasília, DF, Brazil.
[3]General Coordination of Arbovirus Surveillance of the Ministry of Health, Brasília, DF, Brazil.

*Corresponding author(s). E-mail(s): eduardo.krempser@fiocruz.br;
Contributing authors: jaqueline.angelo@fiocruz.br; livia.abdalla@fiocruz.br; daa@fiocruz.br; luisa.spolavori@fiocruz.br; gerlania.rodrigues@aluno.uepb.edu.br; alessandro.romano@saude.gov.br; daniel.ramos@saude.gov.br; pedro.passos@saude.gov.br; marcia.chame@fiocruz.br;

Abstract

Purpose: This study present the evolution of the alert generation framework of the Brazilian Wildlife Health Information System (SISS-Geo), a citizen science platform supporting wildlife health surveillance and zoonosis prevention.

Methods: Three complementary alert strategies were investigated: (i) a rule-based approach based on predefined expert criteria, (ii) an index-based approach that combines multiple cluster attributes through multi-objective optimization, and (iii) an anomaly-based approach that employs unsupervised machine learning to identify atypical spatiotemporal clusters. The methodologies were evaluated using two real-world case studies. The first focused on Yellow Fever surveillance using non-human primate records and official data from the Brazilian Ministry of Health, while the second addressed Avian Influenza surveillance using bird records and outbreak data from the World Organisation for Animal Health (WOAH).

Results: The findings indicate that the three strategies offer complementary capabilities for identifying records and clusters of interest, supporting risk assessment, facilitating detection and early warning in potential health and biodiversity emergencies.
Conclusion: Integrating these strategies can strengthen wildlife disease surveillance by supporting the prioritization of field investigations and resource allocation. The framework contributes to scalable, data-driven decision support within the One Health approach.

Keywords: Early warning system, disease surveillance, spatiotemporal clustering, index-based, anomaly detection

1 Introduction

Early alert systems for zoonotic diseases play a critical role in enabling timely detection and response to emerging outbreaks, thereby reducing detection delays and strengthening preparedness and response in public health. Zoonoses, defined as infectious diseases naturally transmitted between vertebrate animals and humans, represent a major and growing threat to global health. It is estimated that approximately 60% of emerging infectious diseases (EIDs) are zoonotic, and nearly 72% of these events originate from wildlife reservoirs [1]. This trend is largely driven by pressures resulting from human activities, including land-use change, urban expansion, globalization, and climate change, which intensify interactions at the human–animal–environment interface and foster the emergence of pathogens [2].

In this context, enhancing disease surveillance and reporting mechanisms through innovative, data-driven approaches has become imperative. Integrating digital technologies, analytical methods, and signal detection through community-based surveillance is essential to build more resilient and proactive surveillance systems capable of anticipating EID threats [3]. These efforts increasingly align with the One Health paradigm, which recognizes the intrinsic interdependence among human, animal, and environmental health and promotes coordinated surveillance across these domains [4, 5].

The One Health perspective is especially pertinent in biodiverse countries such as Brazil, where complex ecological dynamics, extensive wildlife—human interactions, and environmental heterogeneity pose significant challenges to zoonotic disease surveillance [6]. Effective early warning systems require the integration of environmental drivers of disease emergence and heterogeneous epidemiological data, demanding flexible, scalable, and adaptable surveillance strategies [7, 8].

The need for innovative approaches to enhance surveillance and reporting mechanisms, integrating technology and community involvement are crucial to build a more resilient and proactive global defense against EIDs threats [3]. These systems are moving toward a “One Health” approach, recognizing that the health of humans, animals, and the environment are interconnected [4, 5]. This integrated perspective is particularly relevant in Brazil, where the interplay between environmental and epidemiological factors as well as the reliance on health data demands comprehensive and adaptive strategies for disease surveillance systems [7, 8].

Despite recent advances, significant challenges remain in designing and implementing wildlife and public health surveillance systems that effectively capture relevant data across

diverse ecological and socio-political contexts [9, 10]. Key knowledge gaps include limited methodologies of data collection and systems integration, low quality, inconsistent and fragmented data, limited laboratory infrastructure, and a lack of adoption of emerging technologies such as artificial intelligence and mobile platforms [11, 12]. These gaps hinder the capacity to detect zoonotic pathogens early and to understand transmission dynamics comprehensively, with consequences including delayed outbreak responses and increased public health risks [13], contributing to higher morbidity and mortality from zoonotic diseases in Brazil.

Global public health strategies have emphasized the development of structured systems for the early detection and response to health threats. Early Warning and Response (EWAR) systems are designed to support timely action in situations that require analysis, risk assessment, and monitoring of responses to acute public health events and emergencies [14]. In Brazil, EWAR activities are coordinated through the integration of Indicator-Based Surveillance (IBS) and Event-Based Surveillance (EBS), supported by the National Network of Strategic Health Surveillance Information Centers (CIEVS). The IBS relies on the systematic collection and analysis of structured data from formal health information systems, while EBS focuses on the rapid detection and assessment of unstructured, informal information or unconventional data sources, regarding potential public health threats. The CIEVS network, established in accordance with the International Health Regulations (IHR 2005), is a decentralized network of units distributed across states, municipalities, and strategic regions, including international borders, operating continuously to support timely detection and coordinated response to public health emergencies [15]. In Brazil, CIEVS are present across all Brazilian states, with a total of 190 units located in national and regional reference laboratories, border regions, Indigenous districts, and other strategic areas.

Within the Brazilian EBS framework, surveillance activities include the monitoring of rumors, unofficial reports, media sources, email notifications, and other informal communication channels that may indicate unusual epidemiological events or signs of potential public health emergencies. In this context, participatory platforms such as the Brazilian Wildlife Health Information System (SISS-Geo) contribute to the EBS ecosystem by incorporating citizen-generated wildlife observations as potential signals of zoonotic threats, complementing traditional surveillance mechanisms based on formal health notifications and other EBS strategies.

SISS-Geo [16, 17] is a Brazilian platform developed by the Institutional Platform for Biodiversity and Wildlife Health (Pibss) at the Oswaldo Cruz Foundation (Fiocruz) to monitor the health of wild animals in their natural environments. As a citizen science initiative, it supports both the prevention and control of zoonoses and the conservation of biodiversity. The records of both dead and alive animals support EBS investigations into causes of death and potential human influence, while also provide data for monitoring population dynamics. Its applications are multidimensional: beyond supporting the investigation of events potentially associated with infectious agents, SISS-Geo also supports the monitoring of pathogens of public health significance and wildlife movements across ecological corridors, contributing to the forecasting of disease transmission scenarios, the anticipation of prevention and control measures, and the conservation and preservation of ecosystems.

Building on this foundation, the alert system within SISS-Geo has been evolving through successive methodological stages to enhance its ability to detect potential zoonotic threats.

The current stage relies on fixed rules, designed to identify risk situations based on predefined rules and expert-defined conditions. Although this rule-based approach provided a clear and transparent framework for initial analyses, it was limited in flexibility and unable to represent the complex spatial–temporal dynamics that characterize wildlife health data.

To address these limitations, the proposed framework advances toward two data-driven approaches for alert generation. The first one is an attribute-based index, referred to as the Z-Alert [18]. This index combines multiple attributes derived from spatiotemporal clusters to quantify the level of attention associated with each cluster. By weighting and integrating multiple cluster attributes, the Z-Alert enables finer risk assessments and customizable alert criteria for specific epidemiological or ecological contexts. This transition from fixed rules to a continuous index of attention expanded the analytical capacity of the platform, supporting a more adaptive and evidence-based decision process.

In the second approach, the alert framework incorporates unsupervised anomaly detection techniques to identify atypical patterns directly from the data [19]. This data-driven approach operates without predefined labels or expert rules, allowing the detection of unusual events that may indicate emerging threats. By integrating spatial and temporal dimensions with machine learning-based anomaly detection, SISS-Geo advances toward a more autonomous and adaptive tool capable of supporting EWAR in the context of zoonotic disease monitoring.

Taken together, these three methodological stages (rule-based, index-based and anomaly-based) represent a progressive evolution in the design of alert mechanisms within SISS-Geo, transitioning from static rule-based detection to adaptive, data-driven intelligence. Each approach contributes distinct and complementary advances to wildlife health surveillance. The proposed methodology aims to support wildlife health surveillance by enabling proactive identification of critical regions and efficient resource prioritization. The main contributions of this study are as follows:

- The integration of complementary alert strategies within a coherent analytical framework in SISS-Geo, combining fixed rules, the attribute-based Z-Alert index, and unsupervised anomaly detection to enhance early detection, support decision-making, and improve responsiveness to emerging zoonotic threats.
- The comparative evaluation of alert methodologies, assessing their performance, interpretability, and limitations to characterize their respective strengths, limitations, and complementary roles in identifying areas of concern.
- The advancement of participatory surveillance, reinforcing the role of SISS-Geo as a citizen science platform capable of supporting community-based surveillance strategies by integrating expert knowledge, environmental data, and machine learning for proactive wildlife health monitoring.

Together, these contributions establish a conceptual and methodological foundation for the integration of traditional, unconventional, and data-driven surveillance strategies. This work is structured to facilitate the incorporation of surveillance data, where animal-related reports are systematically processed to generate different types of alerts. These alerts support surveillance activities contributing to the EWAR process. The following sections present related works, followed by the description of SISS-Geo, materials and methods employed, computational results obtained, and conclusions.

2 Related works

A survey of PubMed, SciELO, Elsevier, Medline, WHO LIS, and the Cochrane Library identified varying levels of alignment with SISS-Geo's alert system, particularly in the use of territorial information for health surveillance.

First, we conducted a search in the aforementioned databases using the descriptors "early detection of zoonoses through smart alerts", "early detection of zoonoses by artificial intelligence", and their Portuguese equivalents, within the time frame from 2005 to 2025. We then carried out an initial screening of abstracts to select studies that were closest to SISS-Geo's scope. Among the selected works, we highlight "The evolution of Kenya's animal health surveillance system and its potential for efficient detection of zoonoses" [20], which directly relates to the SISS-Geo proposal by emphasizing the importance of structured surveillance systems, intersectoral integration, and the One Health approach. Although its focus is institutional and oriented toward animal-health surveillance, it presents principles such as systematic data organization, risk-based surveillance, and support for decision-making. In this review, the authors identified 21 relevant papers and highlighted the use of Epicollect, a mobile technology that allows the creation of forms for data collection, analysis, and feedback, with data stored on the server of Kenya's Directorate of Veterinary Services (DVS). In addition to Epicollect, the study discusses mDharura, a system designed to support event-based surveillance in which signals from both sectors are collected by community members and shared across sectors, both implemented within the Kenya Animal Biosurveillance System (KABS) and the Kenya Health Information System (KHIS). This is therefore similar to the set of tools that make up SISS-Geo, whether through direct public participation (participatory surveillance) or through the generation of event-based alerts via automatic email notifications that the platform sends to health authorities based on the analysis of received reports. These tools are used by veterinarians, researchers, forest rangers, and other professionals who interact with sick wildlife.

Another study aligned with SISS-Geo is "Modern technologies and solutions to enhance surveillance and response systems for emerging zoonotic diseases" [3], as it advocates the use of geoprocessing, geographic information systems, and the integration of environmental, epidemiological, and territorial data. The study provides a theoretical basis for using georeferenced tools to support planning, prioritization of actions, and health risk management. In this review, the authors provide valuable information on how technological innovations can be leveraged to prevent, detect, and control outbreaks of emerging zoonotic diseases, and they point to several technologies and computational solutions that can improve surveillance and response systems (such as artificial intelligence and machine learning), citing the BlueDot AI algorithm which, according to the authors, provided an early alert for the COVID-19 outbreak by analyzing news reports.

The paper "Early warning diagnostics for emerging infectious diseases in developing into Late-Stage Pandemics" [21] presents a chronological literature review of various early-detection technologies according to disease progression, from the pre-human phase (prediction and monitoring in animals) to the late stages of pandemics (large-scale rapid diagnostics). It emphasizes the need for early warning systems for emerging diseases and shows that detection delays compromise public-health response. Although its focus is on diagnostics and laboratory technologies, the proposal indirectly reinforces the role of systems such as SISS-Geo, which can support anticipatory surveillance strategies by identifying

priority territories for intensified monitoring. The article also cites BlueDot's AI algorithm as a promising emerging method for identifying diseases with outbreak potential, but concludes that there is no single technology capable of preventing future pandemics; rather, an integrated, multi-phase surveillance system is needed, combining environmental prediction, animal monitoring, population-level analysis (wastewater and air), and rapid diagnostics in humans.

Conversely, "Transmissibility of emerging viral zoonoses" [22] makes a relevant contribution to biological-risk prioritization through machine-learning models. However, its application is predominantly biological and predictive, with no direct interface with territorial management or epidemiological surveillance systems based on georeferenced information, which limits its practical adherence to the scope of SISS-Geo. In that study, a machine-learning predictive model was used to identify specific viral species with potential for human-to-human transmission. The analysis employed gradient-boosted regression trees, and the normalized relative influence of each predictor variable across the 40 tree models was calculated using Friedman's algorithm (the modeling framework was based on published data/observed viral characteristics—genome segmentation, lipid envelope, among other attributes).

Complementarily, the review "Innovative applications of artificial intelligence in zoonotic disease management" [23] broadens the discussion by presenting AI applications in spatial modeling, outbreak prediction, and integrated data analysis, specifically in epidemiological diseases. The article points to possible pathways for the evolution of such systems by incorporating more advanced analytical layers without undermining their role as a structural basis for territorial surveillance. The review discusses contact tracing (in short, its goal is to identify and monitor individuals who may have been in contact with an infected person, thereby preventing transmission; AI methods can help improve the process and speed it up), using the COVID-19 pandemic as an example and highlighting the challenge of identifying asymptomatic cases. When an individual is diagnosed with COVID-19, the application promptly sends automated and anonymous risk notifications to close contacts, accompanied by a request to self-isolate. The rapid issuance of contact alerts and isolation recommendations reduced the spread of the pandemic. Another system mentioned in the same study is COVICT, proposed by Wahid et al., which is a COVID-19 detection and monitoring system based on the Internet of Things (IoT). It uses real-time personal symptom data and contact tracing, employing device-to-device (D2D) communication and gradient-boosting algorithms to determine proximity and contact duration. In addition, the authors argue that the effectiveness of AI-based contact-tracing solutions depends on their widespread adoption and use.

Finally, the study "Risk predictive models of healthcare-seeking delay among imported malaria patients" [24] presents a relevant use of AI in the context of clinical epidemiological surveillance and access to health services. However, its contribution to the scope of SISS-Geo is limited because it does not address the territorial dimension of health surveillance nor the planning of regulatory actions. Although it is a relevant study, it focuses heavily on the risk of delayed care seeking among imported malaria patients in Jiangsu Province, and the authors point out several limitations: the exclusive use of treatment within 24 hours as the criterion for delayed treatment, limited sample size, and the lack of external validation of the constructed model.

The analyzed papers show that the progressive incorporation of AI together with advanced analytical techniques can enhance georeferenced information systems, strengthening surveillance as a strategic instrument for preventing and protecting public health. Early detection of zoonotic diseases can help control outbreak spread and, consequently, reduce the damage caused. Despite these advances, none of the reviewed studies integrates citizen science, wildlife and public health surveillance, spatiotemporal clustering, alert prioritization, anomaly detection, and integration with official health surveillance systems into a unified framework.

3 SISS-Geo

The Wildlife Health Information System (SISS-Geo) is a free platform available for smartphones and on the web[1]. It enables the monitoring of wildlife health across natural, rural, and urban environments, supporting the investigation of disease-causing agents that can affect both humans and animals. As a citizen science initiative, it engages ordinary citizens, health and environmental professionals, researchers, and specialists to contribute by providing records, fostering actions for zoonosis prevention, disease control, and biodiversity conservation in Brazil.

Since its launch in March 2014, SISS-Geo has been downloaded more than 10,000 times from the Google Play Store. By March 2026, it had accumulated over 65,000 records worldwide, approximately 32,000 of which were taxonomically validated by experts, with contributions from more than 20,000 users, including citizens, specialists, and institutional partners. Figure 1 presents the geographic distribution of georeferenced observations recorded by SISS-Geo until March 2026. Brazil accounts for the highest number of records, but entries have been registered in over 20 other countries.

SISS-Geo contributes to wildlife health surveillance and zoonosis prevention in multiple ways. It begins by generating real-time alerts whenever abnormalities in wild animals are detected. In this scheme, alerts are automatically triggered when predefined conditions are met. For example, whenever a report of a sick or dead non-human primate (NHP) is submitted to the platform, an email notification is immediately sent to registered health managers, including the details of the occurrence. This mechanism is simple and ensures rapid communication of potentially zoonotic events. In practice, this means that dead or diseased NHPs must be located and reported so that biological samples can be collected and sent for diagnosis as quickly as possible, as carcass decomposition can obscure important lesions and compromise the effectiveness of surveillance.

The system also enhances the flow of information across different sectors. By connecting citizens, professionals, decision-makers, and researchers, it ensures that knowledge circulates quickly and efficiently, enabling the identification of priority areas and the implementation of integrated responses. Monitoring is another essential contribution. With the active participation of the public, the system collects photos and georeferenced data of wildlife observations and local features. Experts validate these records, identify the species involved, and provide the resulting information to health services, environmental agencies, research institutions, and society. As such, SISS-Geo's curated records and alerts further enhance the measures against

[1]The system is accessible at https://sissgeo.fiocruz.br

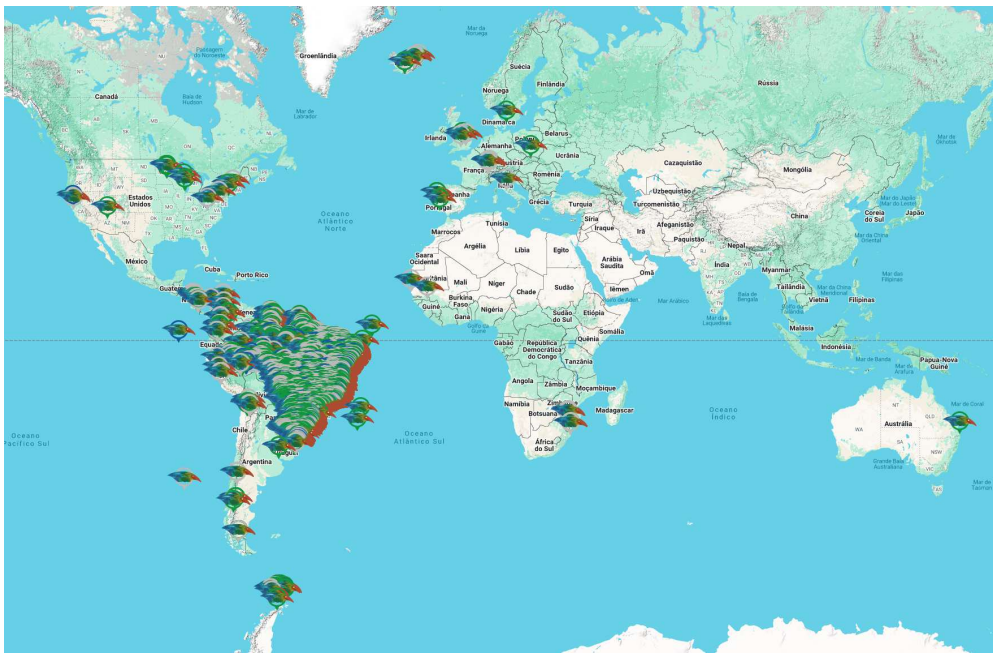


Fig. 1: Geographic distribution of SISS-Geo's records around the world until March 2026. *Source: SISS-Geo's Map (https://sissgeo.fiocruz.br/mapaRegistrosInicial.xhtml)*

the (re)emergence of zoonoses, supporting strategies and public policies focused on prevention, species management, and biodiversity conservation in Brazil. Table 1 presents the fifteen most frequently reported animal types in SISS-Geo up to March 2026.

SISS-Geo has become an essential resource for supporting decision-making during major wildlife health events in Brazil, such as Yellow Fever (YF) outbreaks [25–27]. By accelerating information dissemination, providing precise geographic data, and facilitating coordinated response efforts, the platform has significantly strengthened national surveillance capacity. Building upon this trajectory, a technical note issued by the Brazilian Ministry of Health (BMoH) in 2025 [28] formally adopted the use of SISS-Geo for registering all reports of sick or dead NHP for epidemiological surveillance purposes. This official endorsement consolidates the system's position as an institutional tool for early detection and rapid response to YF and other zoonotic threats in Brazil.

Tools such as SISS-Geo and the environmental suitability models developed by Fiocruz Institutional Biodiversity and Health Platform (Pibss) in partnership with the BMoH [29], play an increasingly important role in anticipating areas of elevated risk, informing vaccination and surveillance strategies, and supporting the identification of ecological transmission corridors and enabling more accurate risk prediction. In this context, ongoing efforts to strengthen surveillance strategies focus on the progressive improvement of the alert system, through the integration of regional information, aiming to enhance the accurate identification of risk events and support timely public health responses.

Table 1: Fifteen most frequently reported animal types in SISS-Geo until March 2026.

Animal: Type	Number of records	%
Bird	8626	15.38
Bat	6417	11.44
Monkey	5507	9.82
Marmoset	5178	9.23
Bird: Other	3574	6.37
Other	3168	5.65
Snake/Serpent	2204	3.93
Bird: Bird of Prey	2063	3.68
Bird: Aquatic/Marine	2048	3.65
Bird: Pigeon	1699	3.03
Opossum	1628	2.90
Amphibian	1566	2.79
Bird: Macaw/Parrot/Parakeet	1201	2.14
Dog	1199	2.14
Capybara	931	1.66

4 Materials and Methods

This study explores three approaches for generating early warning alerts within the SISS-Geo platform. Each method reflects a distinct stage in the evolution of surveillance systems, ranging from simple rule-based mechanisms to more sophisticated, data-driven techniques. At present, only the rule-based alert system is operational and routinely used for surveillance activities. The index-based and anomaly-based approaches have reached research maturity and are currently undergoing deployment within the platform to improve alert-generation capabilities. Figure 2 illustrates the conceptual differences between the three alert generation strategies explored in this study. Each approach represents a distinct mechanism for transforming wildlife observation records into epidemiological alerts, differing in their level of automation, reliance on prior knowledge, and analytical complexity.

The rule-based approach corresponds to the most straightforward strategy for generating alerts. In this case, notifications are triggered whenever predefined conditions are satisfied, such as reports of NHP deaths, sick animals, or confirmed cases of diseases of interest. Other types of alerts may include the appearance of a specific animal species, threatened with extinction, or any animal presence in designated areas. These rules are defined by domain experts and reflect established surveillance protocols by the Brazilian government, commonly used in public health systems. Although this approach is transparent, easy to implement, and is operationally useful for rapidly mobilizing the local teams responsible for collecting biological samples and the subsequent steps of the investigation, it is inherently limited by its dependence on explicitly defined criteria, which may fail to capture emerging patterns that do not fit predefined rules. In contrast, index- and anomaly-based approaches offer more flexibility for detecting novel behaviors.

The index-based approach extends this logic by combining multiple attributes from a set of records (clusters) into a single quantitative measure, indicating the level of attention associated with each cluster. Following spatial-temporal clustering of wildlife records, several

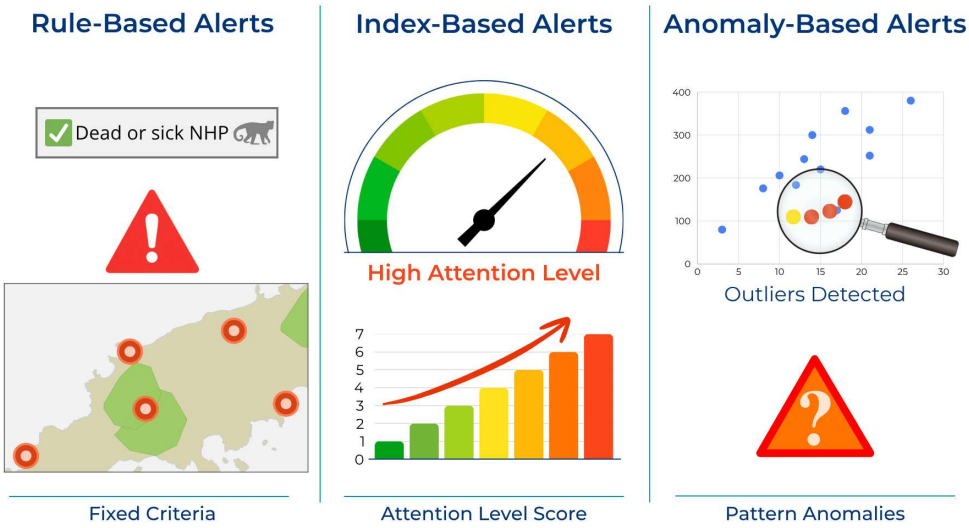


Fig. 2: Conceptual illustration of the three alert generation strategies explored in this study. The illustration was generated using Canvas.

cluster-level attributes are extracted and integrated through a weighted aggregation process. The relative importance of these attributes is calibrated using a multi-objective optimization technique, allowing the alert index to better reflect the defined optimization objectives. Clusters with higher index values represent situations with greater attention and are therefore recommended to be prioritized for analysis and response by surveillance authorities [18].

Finally, the anomaly-based approach adopts a data-driven perspective by focusing on the detection of unusual patterns rather than predefined signals. After the clustering and feature extraction stages, unsupervised machine learning models are applied to identify clusters that deviate from typical patterns observed in the dataset. Methods such as One-Class Support Vector Machine, Isolation Forest, and Local Outlier Factor estimate anomaly scores that indicate how atypical a cluster is relative to the overall distribution of observations. Clusters identified as anomalies are interpreted as potential early warnings of emerging epidemiological events [19].

Together, these three approaches form a complementary framework for alert generation. While rule-based alerts ensure immediate detection of well-known epidemiological signals, the index-based system enables systematic prioritization of clusters based on multiple attributes, and the anomaly detection models expand surveillance capacity by revealing unexpected or unknown patterns.

4.1 Rule-based Alert System

The rule-based approach constitutes the initial stage of the SISS-Geo alert system. In this scheme, alerts are automatically triggered whenever predefined conditions are met, and an email notification is automatically sent to registered managers, with details of the occurrence.

Figure 3 presents examples of email notifications sent to health managers, including details of the reported record and the location of the event.

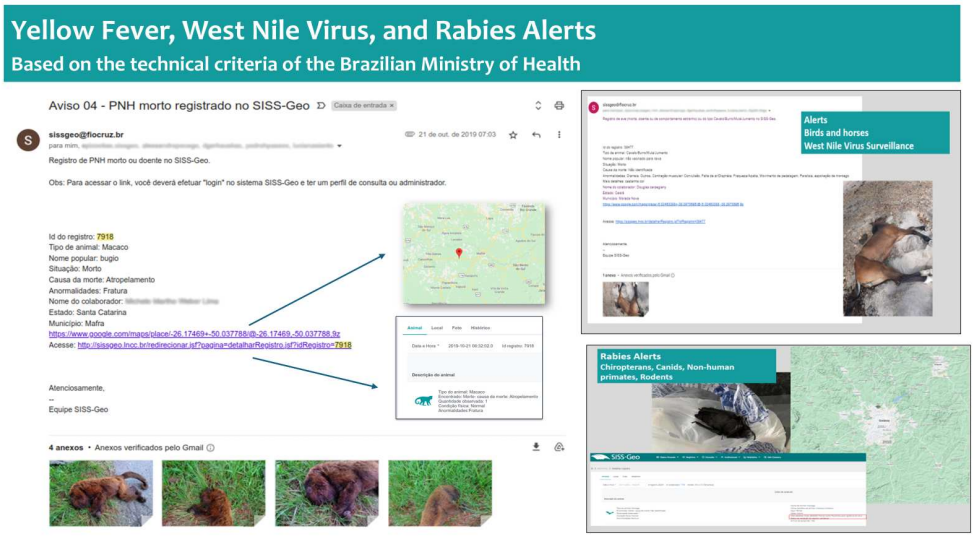


Fig. 3: Alert generated by the rule-based system, providing information to health manager. *Source: Institutional Platform for Biodiversity and Wildlife Health (Pibss)/Fiocruz.*

The rule-based alert system relies on predefined rules that evaluate the characteristics of each record submitted to the SISS-Geo platform. Each rule represents a specific condition of interest and is designed to identify records that meet predefined epidemiological or surveillance criteria. When a new record satisfies the conditions associated with a rule, the corresponding alert is triggered. A single record may satisfy multiple rules simultaneously, resulting in the activation of multiple alerts.

The relationship between records, rules, and email notifications is therefore not one-to-one. A given record can trigger several rules, while each triggered rule can generate alerts to multiple recipients. This structure allows the same observation to be communicated simultaneously to different managers, health professionals, specialists, or other stakeholders according to the purpose and scope of each rule. Figure 4 presents a scheme of the current rule-based alert system generation.

For example, the registration of a dead bat may satisfy more than one surveillance criterion. The same record may trigger the rule “Alert: Bat, anteater, armadillo, or sloth recorded in SISS-Geo” and the rule “Alert: Dead or sick bat in the State of Goiás”. Each of these rules may, in turn, be associated with multiple email recipients. Consequently, a single record can activate multiple rules and can trigger several email notifications, potentially reaching different stakeholders with an interest in the event.

Despite the SISS-Geo system has been generating alerts since 2019, their systematic monitoring began in April 2024. Between April 2024 and July 2026, SISS-Geo had 334 active rules. In this period, 48,473 alerts were triggered by 16,055 records, resulting in 114,738

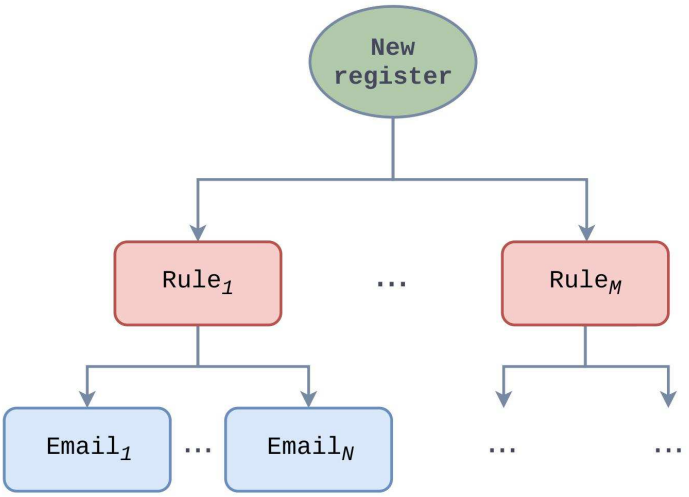


Fig. 4: Scheme of the rule-based alert system.

emails sent. Figure 5 shows the number of emails triggered for each animal type, while Figure 6 presents the number of records that activated at least one rule for each animal type. It is interesting to note that, although a larger number of bat-related records activated rules, the number of emails triggered for monkeys and marmosets is considerably higher than that associated with bat-related records. This difference can be explained by Brazil's long-standing history of YF surveillance, which has led to more intensive monitoring of NHPs. Particularly in the context of YF surveillance, all reports of sick or dead NHPs registered in SISS-Geo must be notified to health authorities. As a result, records involving NHPs can generate a higher number of email notifications, reflecting the greater relevance of these observations to ongoing disease surveillance. This high volume of email notifications raises the question of whether all alerts can be effectively investigated.

Although effective for generating basic alerts and signaling potential risks–including the possible detection of exotic species–this system provides limited prioritization among triggering alerts and can generate a large volume of notifications. Because all triggering alerts are treated equivalently, the system does not distinguish their relative level of importance, which may place a substantial overload on surveillance teams. Consequently, some alerts may receive delayed attention or remain unassessed, highlighting the need for more selective strategies to support the prioritization and efficient allocation of surveillance resources, particularly before the confirmation of zoonotic pathogens.

Nevertheless, these alerts play a fundamental role in surveillance, as they direct field investigations and the collection of biological material, which provide the basis for generating subsequent alerts. This feature is particularly important given the integrated and hierarchical network of the Brazilian Unified Health System (SUS), especially during large-scale events

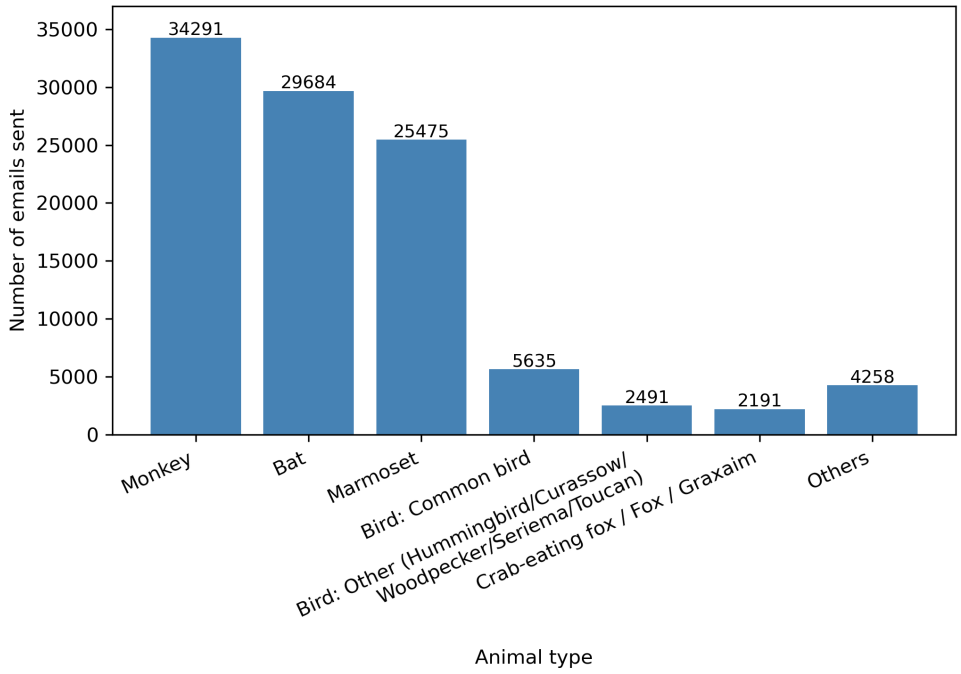


Fig. 5: Emails sent by SISS-Geo system per animal type.

affecting multiple territories, as it supports decision-making by government entities regarding the responses to significant health and biodiversity conservation events at the municipal, regional, and national levels.

To leverage data from the SISS-Geo platform to improve surveillance strategies, two complementary data-driven frameworks were proposed: the index-based and the anomaly-based models. Those models are designed to incorporate multiple attributes, capture complex patterns in the data, and generate more prioritized alerts to support decision-making and the sensitivity of epidemiological surveillance in detecting signs of potential emergencies involving biodiversity and public health.

4.2 Data-driven Alert Systems

The index-based and the anomaly-based models rely on a common initial pipeline that ensures data consistency and enables the identification of meaningful epidemiological patterns. This pipeline comprises two essential steps: (i) collection and preprocessing of records from the SISS-Geo platform, and (ii) performing spatial-temporal clustering on the preprocessed records to generate cluster-level attributes, thereby enabling the detection of health-related phenomena. These steps serve as the basis upon which each methodology builds its alert-generation logic, enabling more robust and interpretable epidemiological insights. The following subsections describe these steps in detail.

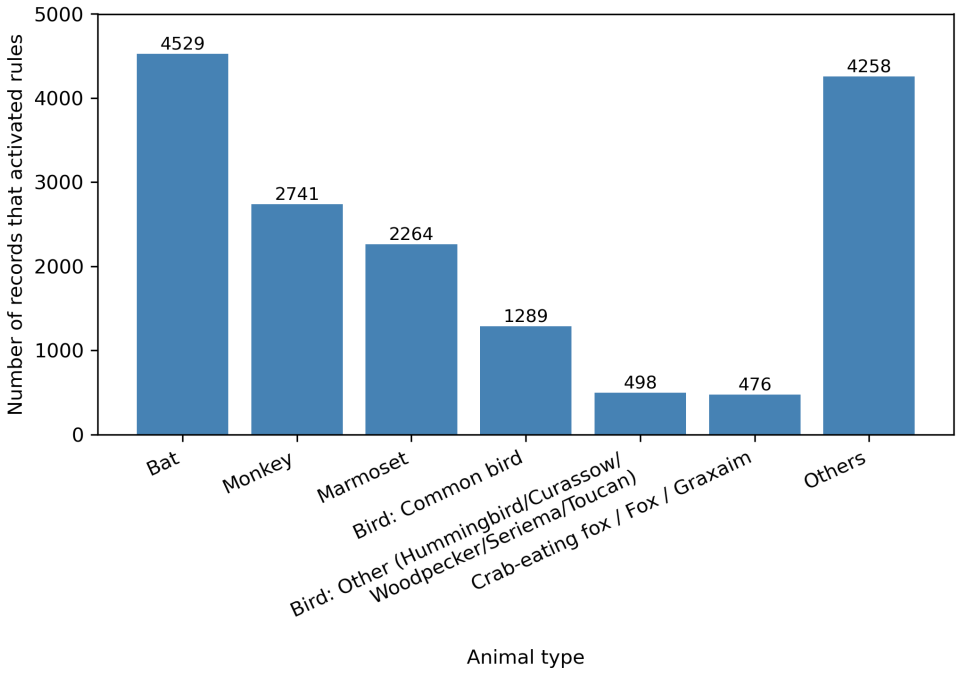


Fig. 6: Records that triggered alerts per animal type.

4.2.1 Data preprocessing and Clustering

A. Data collection

For both models, the first step involves the collection and preprocessing of SISS-Geo's data. This step is essential for clustering the records and extracting cluster attributes to support subsequent analyses.

Table 2 presents the data collected from the SISS-Geo platform. The selection of these attributes was guided by their relevance to the data-driven approaches. Spatial descriptors (longitude, latitude, location source, and precision) enable the structuring of spatial clusters and the identification of geographic patterns. Temporal and contextual record information supports the characterization of event dynamics, while record status attributes ensure data reliability and quality control. Animal-related fields (type, observed quantity, situation, behavior, physical condition, and cause of death) provide biologically and epidemiologically meaningful signals associated with abnormal morbidity and mortality patterns. Finally, the final classification fields (disease and classification) offer a validated reference layer for validation, and performance assessment. Together, these attributes support the development of a generalizable alert index and the detection of statistically and epidemiologically relevant anomalies across different zoonotic and environmental health surveillance contexts.

Although the public SISS-Geo interface provides access to a limited subset of the data, such as location, date, and animal species, the full dataset used in this study contains additional information, including disease outcomes, and confirmed cases, accessible only to

Table 2: Data collected from SISS-Geo's platform.

Field	Descriptor field
Record	Identifier, Longitude, Latitude, Observation date, State, Municipality, Source of location and Precision.
Record status	"Checking with experts" (records verified by the SISS-Geo internal team and awaiting validation by a taxonomist), "Audited" (records verified by the SISS-Geo internal team), and "Valid" (records validated and approved by taxonomist).
Animal	Identifier, Type, Observed quantity, Animal condition (alive or dead), and Behavior (if alive, normal or sick).
Final classification	Disease (disease under investigation) and Classification (positive or negative for the tested disease).

authorized institutional users due to legal and privacy constraints.

B. Spatiotemporal Clustering

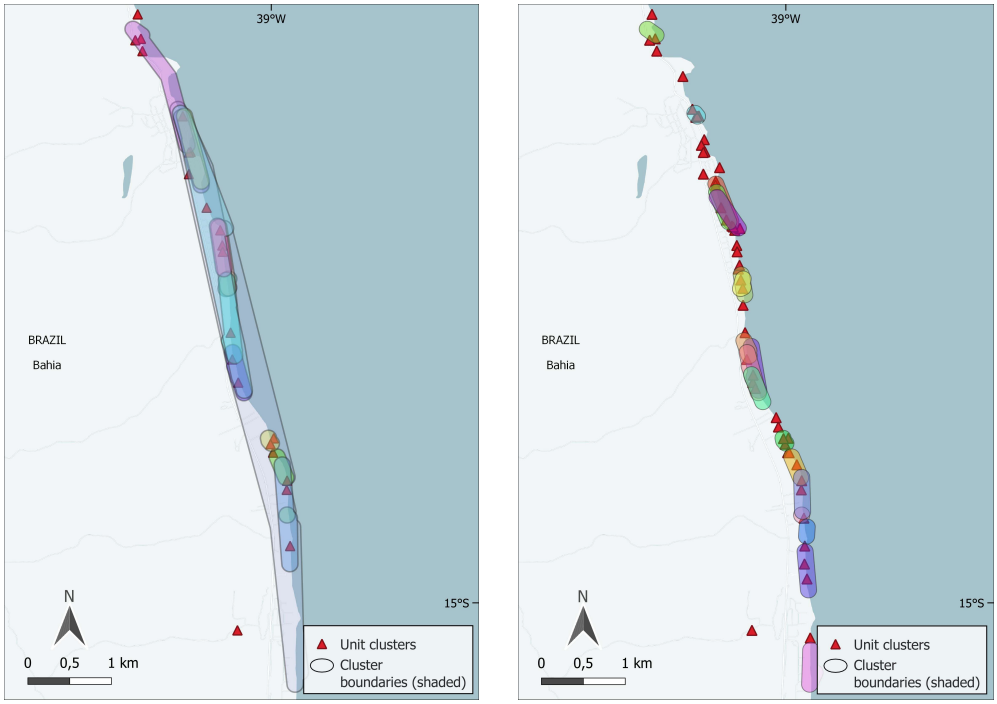
The clustering stage was applied in both the index-based and anomaly-based frameworks. Clustering techniques group data with shared characteristics, enabling the identification of instances that likely belong to the same phenomenon. For dynamic processes, such as disease outbreaks, records often exhibit spatial and temporal patterns [27]. Clustering therefore helps to identify groups of records that follow similar dynamics, supporting a more refined analysis of disease spread.

Clustering of SISS-Geo records was performed using the DBSCAN algorithm [30, 31], due to its ability to identify clusters of arbitrary shape, handle noise, and the relatively small number of parameters required for configuration. A spatio-temporal formulation with a pre-computed distance matrix was adopted, combining spatial and temporal components under user-defined constraints. It allows DBSCAN to identify meaningful clusters that reflect the dynamics of disease occurrence while preserving the integrity of isolated but relevant records. Detailed descriptions of construction and normalization of the spatial and temporal distance matrices are provided in [18, 19].

To illustrate the clustering stage, Figure 7 shows two clusters generated from turtle records in SISS-Geo platform using different spatial-temporal parameters. The selected parameter values were chosen solely to illustrate the behavior of the clustering algorithm under different configurations and should not be interpreted as biologically or epidemiologically meaningful thresholds. In 7a clusters were obtained with a maximum spatial distance of 1 km and temporal limit of 30 days, while in 7b a distance of 500 m with temporal limit of 10 days were used as threshold. Each polygon represents a cluster, and red triangles represent unit clusters (noise) corresponding to records that did not meet the minimum density required to form a cluster.

As the spatial and temporal thresholds are reduced, a larger number of records fail to meet the minimum density criteria, resulting in an increased number of unit clusters and more fragmented groupings. This effect reflects a stricter clustering condition, in which only records that are closer in both space and time are grouped. However, it is important to note that these thresholds are applied for pairs of points, meaning that clusters are not strictly bounded by the initial limits. Through a chaining effect, records can be connected via intermediate points, for example, if record A is close to B, and B is close to C, all three may

belong to the same cluster even if A and C are distant in space and time. This property allows DBSCAN to capture the continuity of spatiotemporal patterns, enabling the identification of clusters that reflect the spread or persistence of epidemiological events while maintaining sensitivity to emerging or isolated signals.



(a) Using a 1 km spatial and 30-day temporal threshold.

(b) Using a 500 m spatial and 10-day temporal threshold.

Fig. 7: Illustrative examples of clusters formed by DBSCAN.

C. Cluster characterization

After the clustering step, each cluster is enriched with complementary information that serves as the foundation for building the data-driven alerts. The set of extracted attributes, summarized in Table 3, captures different perspectives of the phenomenon under study, including epidemiological patterns (e.g., number of dead animals, frequency of reports), temporal dynamics (e.g., how concentrated or dispersed the events are in time), and spatial characteristics (e.g., the geographic extension of the cluster). By combining these attributes, the characterization process transforms raw groupings of records into structured units of analysis that can be meaningfully compared.

After this enrichment stage, the clusters become suitable for further analysis using the two complementary data-driven methodologies. The index-based model relies on the extracted

Table 3: New attributes extracted from clustering.

Attribute	Description
n_death	Number of dead animals.
n_animal	Total number of animals involved; sum of live, sick and dead animals.
ini_date	Cluster start date, corresponding to the oldest record.
end_date	Cluster end date, corresponding to the most recent record.
$interval$	Time interval (in days); difference between the dates of the earliest and most recent records.
n_rec	Number of records.
$freq_rec$	Frequency of records; number of records divided by $interval$.
$freq_death$	Frequency of dead animals; number of dead animals divided by $interval$.
$freq_animal$	Frequency of total animals; number of animals involved divided by $interval$.
$perc_death$	Percentage of dead animals; number of dead animals divided by n_animal.
$extention$	Maximum distance between all points in the cluster.
n_conf	Number of animals confirmed to have a disease of interest.
$geocode$	Cluster geocode.

attributes to construct an alert index, allowing clusters to be ranked according to their relevance, whereas the anomaly-based approach applies machine learning techniques to detect unusual patterns in the data that may indicate early signs of risk. These two techniques are presented in the following subsections.

4.2.2 Index-based Alert System

The Multi-Attribute Zoonotic Alert Index (Z-Alert) [18], represents an advancement over rule-based alerts by integrating clustering analysis with multi-objective optimization. In this approach, the spatial and temporal clustering is first performed and then a set of attributes is extracted, capturing epidemiological, temporal, and spatial dimensions.

To transform these attributes into a single, interpretable measure, a multi-objective optimization procedure is applied. This optimization assigns weights to the attributes in a way that maximizes the mean alert score of critical clusters (clusters with confirmed cases of disease) while minimizing their variance, ensuring both sensitivity and stability of the indicator. The aim is not to optimize a generic mathematical function, but to calibrate the relative importance (weights) of cluster attributes so that the resulting index better reflects epidemiological relevance.

As a result, each cluster is assigned a continuous alert score, the Z-Alert index, which represents its *level of attention*. Rather than predicting the occurrence of a specific disease, the index is designed to prioritize clusters according to their relevance, identifying those that may warrant closer monitoring or further investigation. Its utility is particularly relevant when a high level of attention is assigned to a cluster before disease confirmation is available, as this may support the early prioritization of surveillance and investigative efforts. When the alert follows a confirmed diagnosis, the index can still support the prioritization and characterization of relevant events.

Once the index has been constructed, the next step is to establish how it should be interpreted in practice. The *threshold level* parameter directly affects the number of alerts generated and ensures that only the most relevant situations are reported to the manager. It

establishes a threshold criticality used to determine which clusters generate alerts. Proper calibration of the threshold level reduces the risk of overreacting to low-risk events while still capturing emerging patterns that might not be evident from individual reports.

A detailed description of the methodology is presented in [18], providing a comprehensive view of each step. Figure 8 illustrates such approach. To summarize, the index-based alert system provides health managers with a clear and interpretable way to prioritize alerts associated to clusters. By transforming raw surveillance data into a continuous score and allowing thresholds to be adjusted to local contexts, it ensures that decision-making is both flexible and grounded in epidemiological evidence. This methodology strengthens the ability to focus resources where they are most needed, offering a transparent and systematic framework for public health action.

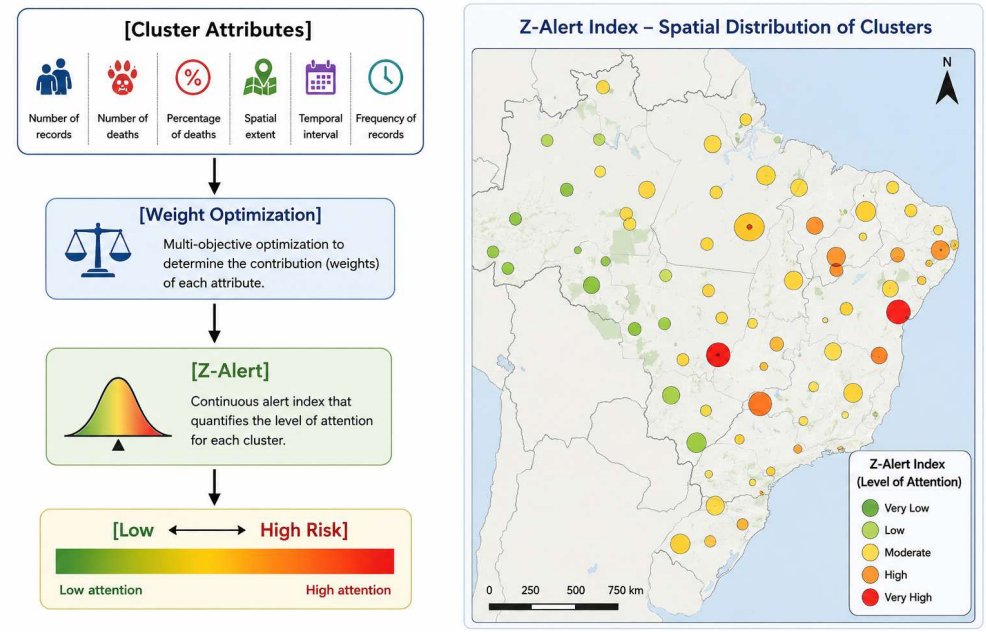


Fig. 8: The Z-Alert Index approach. The illustration was generated using an AI-assisted tool and subsequently refined by the authors.

4.2.3 Anomaly-based Alert System

The anomaly-based approach [19] aims to enhance surveillance by identifying atypical clusters without requiring predefined rules or disease-specific calibration. Following the same spatiotemporal clustering and feature extraction steps used in the index-based approach, unsupervised machine learning models are applied to the resulting cluster-level attributes.

Unlike the index-based approach, which is calibrated using clusters associated with confirmed cases of the disease of interest, anomaly detection does not require disease-specific

reference data to identify unusual patterns. Instead, it learns the distribution of the cluster attributes directly from the available data. This reduces dependence on confirmed cases, which may be scarce or unavailable, and allows the approach to identify atypical patterns that are not necessarily represented in previously characterized disease events.

The number of clusters flagged as anomalous is controlled by the *contamination rate* parameter, which specifies the expected proportion of anomalies in the dataset. This parameter directly affects the number of alerts generated and therefore plays a role analogous to the threshold level in the index-based approach. Its calibration is important for controlling the trade-off between detecting unusual patterns and limiting the number of potentially spurious alerts.

When integrated with the rule-based and the index-based alert systems, the anomaly-based approach provides a complementary data-driven perspective, contributing to a more proactive, flexible and comprehensive public health response. A detailed illustration of the methodology is given in Figure 9.

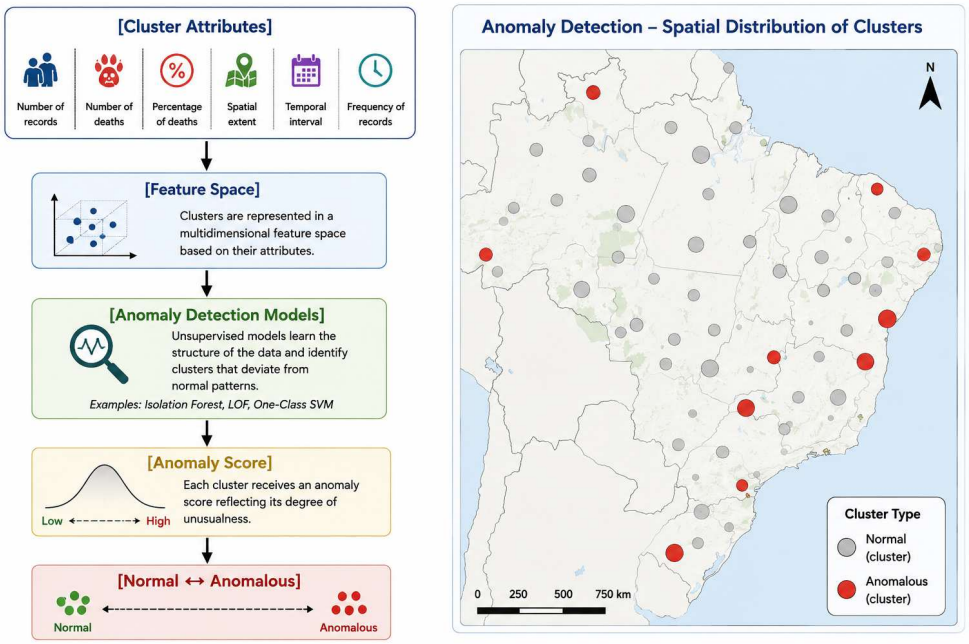


Fig. 9: The Anomaly Detection approach. The illustration was generated using an AI-assisted tool and subsequently refined by the authors.

5 Case Studies

5.1 Epidemiological Scenarios and Datasets

SISS-Geo has evolved into a broader wildlife health surveillance platform, supporting the monitoring of multiple zoonotic diseases nationwide. To evaluate the proposed alert generation strategies, two case studies were conducted using real data collected from the SISS-Geo platform, each addressing a different epidemiological context and animal group.

As SISS-Geo was originally developed to support the surveillance of NHPs in the context of endemic YF in Brazil, the first case study focuses on NHP records as potential signals of YF circulation. In this case study, the rule-based, index-based, and anomaly-based approaches were applied to a set of NHP records. The index-based and anomaly-based methods used the same spatiotemporal clusters and cluster-level attributes, ensuring that their results could be compared under identical spatial and temporal criteria. The results were validated by comparing them with official YF records from the Brazilian Ministry of Health (BMoH) dataset. In addition to evaluating their ability to identify clusters associated with YF-related events, the approaches were compared in terms of the number of alerts generated, providing a complementary assessment of their alerting behavior and degree of selectivity.

The second case study focuses on bird records in the context of avian influenza (AI) surveillance. Unlike the YF scenario, in which SISS-Geo supports the Brazilian public health surveillance system, AI surveillance in Brazil is primarily under the responsibility of the Ministry of Agriculture and Livestock (MAPA), and SISS-Geo is not currently part of the official AI surveillance system. Although SISS-Geo includes fixed-rule alerts for other diseases of interest, such as West Nile virus, it does not currently provide dedicated surveillance support for AI.

As SISS-Geo contains a limited number of confirmed AI cases that could be used to calibrate the Z-Alert index, the index-based alert could not be constructed for this scenario. Instead, only the anomaly-based approach was applied to identify atypical patterns in bird-related observations. This scenario was therefore selected not only to evaluate the anomaly-based approach in a setting with limited disease-confirmed data, but also to explore the potential of SISS-Geo as a complementary source of information for wildlife disease surveillance. In particular, the analysis investigates whether unsupervised models can identify atypical spatial and temporal patterns in bird observations that may be relevant to diseases not currently covered by the platform's surveillance framework.

To provide an external reference for evaluating the anomaly-based alerts, SISS-Geo records were compared with confirmed AI cases reported by the World Organisation for Animal Health (WOAH). The WOAH data were not used for model training; rather, they served as an independent reference to assess whether clusters flagged by the anomaly-based approach showed spatial and temporal correspondence with confirmed AI cases.

In both case studies, the data consist of georeferenced wildlife observations reported by users of the SISS-Geo platform, including information on species, number of individuals, and health condition (e.g., alive, sick, or dead). For the data-driven approaches, the records were preprocessed to ensure data quality and subsequently grouped into clusters using the spatiotemporal clustering procedure, as described in Subsection 4.2.1. These clusters form the basis for the application and comparison of the data-driven alert generation strategies.

5.2 Case Study 1: Yellow Fever surveillance

In these experiments, the approaches were validated using NHP records, specifically monkeys and marmosets, from SISS-Geo covering the period from 01 January 2025 to 31 December 2025.

5.2.1 Comparison of cluster identification strategies

We first analyze the performance of index-based and anomaly-based approaches to evaluate their ability in identifying clusters associated with YF-related events. Data for these analyses were extracted from SISS-Geo on 26 June 2026, considering the fields described in Table 2, resulting in 2,297 wildlife records. After filtering the records to retain only observations acquired via GPS or with coordinates explicitly provided by the user, and with a maximum positional uncertainty of 100 m, 1,401 records remain for analysis.

Using a spatial limit of 1 km and a temporal limit of 30 days, the clustering phase resulted in 1,037 clusters. Among those, 848 were unit clusters, accounting for approximately 81% of the total. Because unit clusters provide limited spatial-temporal context on the analyzed attributes compared to clusters with multiple records, the two approaches were evaluated only for non-unit clusters whose municipalities and event periods matched the corresponding information in the BMoH database. We used a 30-day validation window to confirm an alert. After this filtering, 66 clusters containing 174 records remained for evaluation.

For calibrating the alert index parameters and training the unsupervised anomaly detection models, we used historical data covering the period from 5 May 2014 to 31 December 2024, as described in previous works [18, 19]. The records from 2025 were then used as an independent dataset to assess the performance of both approaches on unseen data.

To assess the performance of the approaches by comparing the generated alerts with the observed occurrence of YF cases, we classified the results into four categories: True Positive (TP), when the system generates an alert and the corresponding municipality has a confirmed YF case within the same period; False Negative (FN), when the system does not generate an alert but the corresponding municipality has a confirmed YF case within the same period; False Positive (FP), when the system generates an alert but the corresponding municipality has no confirmed YF cases within the same period; and True Negative (TN), when the system does not generate an alert and the corresponding municipality also has no confirmed YF cases within the same period.

Based on this classification, Figures 10 and 11 present the performance metrics (precision, sensitivity, specificity, F1 score and accuracy) obtained for the index-based and anomaly-based approaches under different threshold levels *TL* ($0.8, 0.95$, and 0.99) and contamination rate η ($0.5, 0.3$, and 0.1), respectively. We also present the number of alerts (Alert) and detected anomalies (Anom.) for each method. For the index-based approach, we report the solution that provided the best balance among the optimization objectives, and for the anomaly-based approach, we present the results for three anomaly detection methods: One-Class SVM (OCSVM), Isolation Forest (IF), and Local Outlier Factor (LOF).

In Figure 10, the results show a consistent trade-off between sensitivity and specificity as the alert selection criteria become more restrictive. For the index-based approach, increasing the threshold from $TL = 0.8$ to $TL = 0.99$ reduced the number of alerts from 47 to 3

and the number of true positives from 26 to 2. At the same time, the number of false negatives increased from 10 to 34, resulting in a decrease in sensitivity from 0.72 to 0.06. This increase in missed alerts is particularly relevant in wildlife disease surveillance, where failing to detect a condition associated with an emerging event may have consequences not only for wildlife populations and the environment, but also for human health. The intermediate threshold ($TL = 0.95$) provided a different balance, with 21 alerts, the highest precision (0.76), and a specificity of 0.83. The $TL = 0.8$ configuration yielded the highest sensitivity (0.72) and F1-score (0.63), whereas $TL = 0.95$ resulted in the highest accuracy (0.62). These results indicate that the threshold directly influences the selectivity of the alert system, with lower thresholds favoring the identification of a larger proportion of municipalities associated with confirmed YF cases, while higher thresholds substantially reduce the number of alerts and increase the number of false negatives.

A similar pattern was observed in Figure 11 for the anomaly-based approaches as the contamination parameter η decreased. For all three algorithms, reducing η resulted in fewer detected anomalies and higher specificity, but also substantially lower sensitivity. For example, in OCSVM, the number of anomalies decreased from 29 to 4 as η decreased from 0.5 to 0.1, while sensitivity decreased from 0.50 to 0.06 and specificity increased from 0.63 to 0.93. The same trend was observed for IF and LOF. Among the anomaly-based configurations, $\eta = 0.5$ consistently resulted in the highest F1-score for each algorithm, with values of 0.55, 0.63, and 0.58 for OCSVM, IF, and LOF, respectively. These results suggest that TL and η act as mechanisms for controlling the selectivity of the alert generation process. In the context of wildlife disease surveillance, this flexibility may allow the approaches to be configured according to the intended surveillance objective, balancing the identification of potentially relevant events against the volume of alerts requiring subsequent investigation.

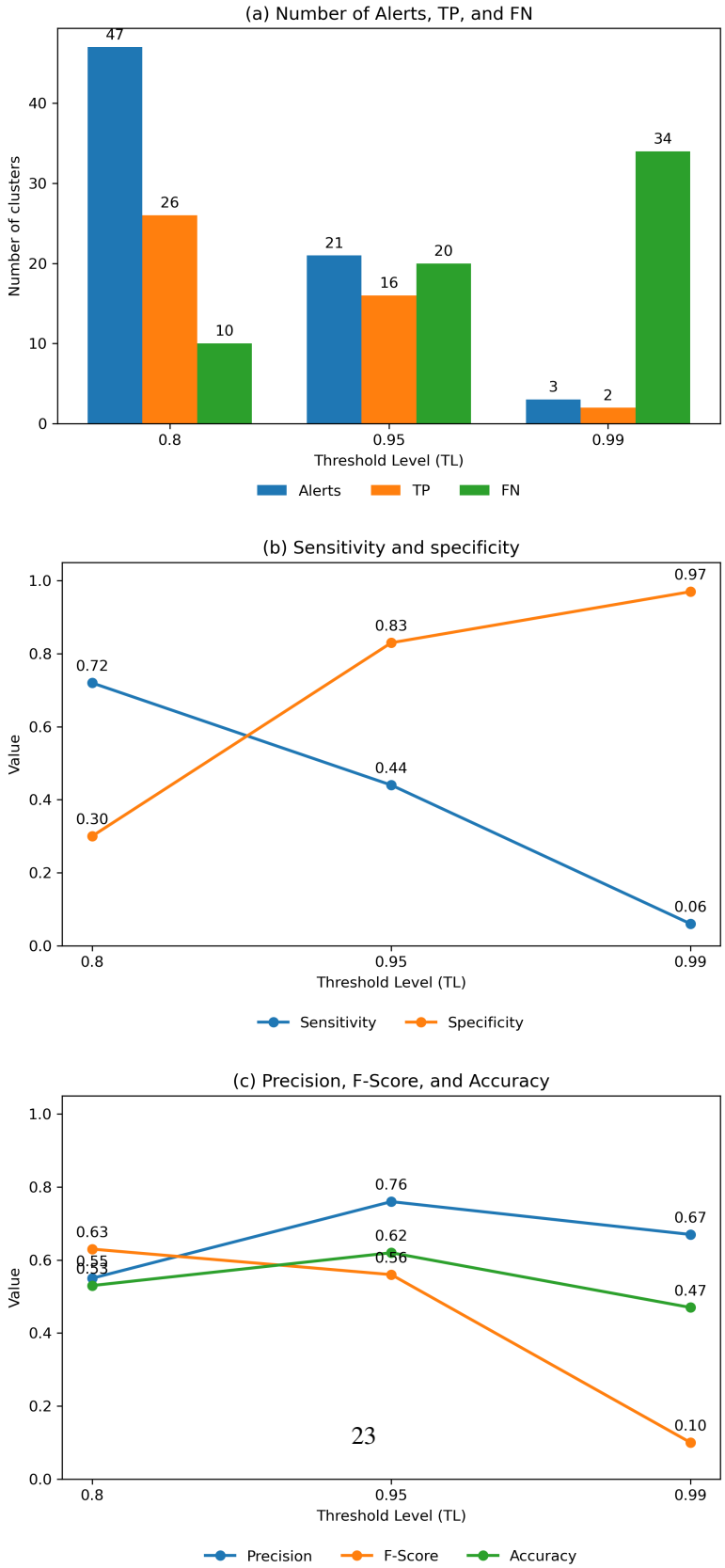


Fig. 10: Performance metrics for index-based approach on data from 2025.

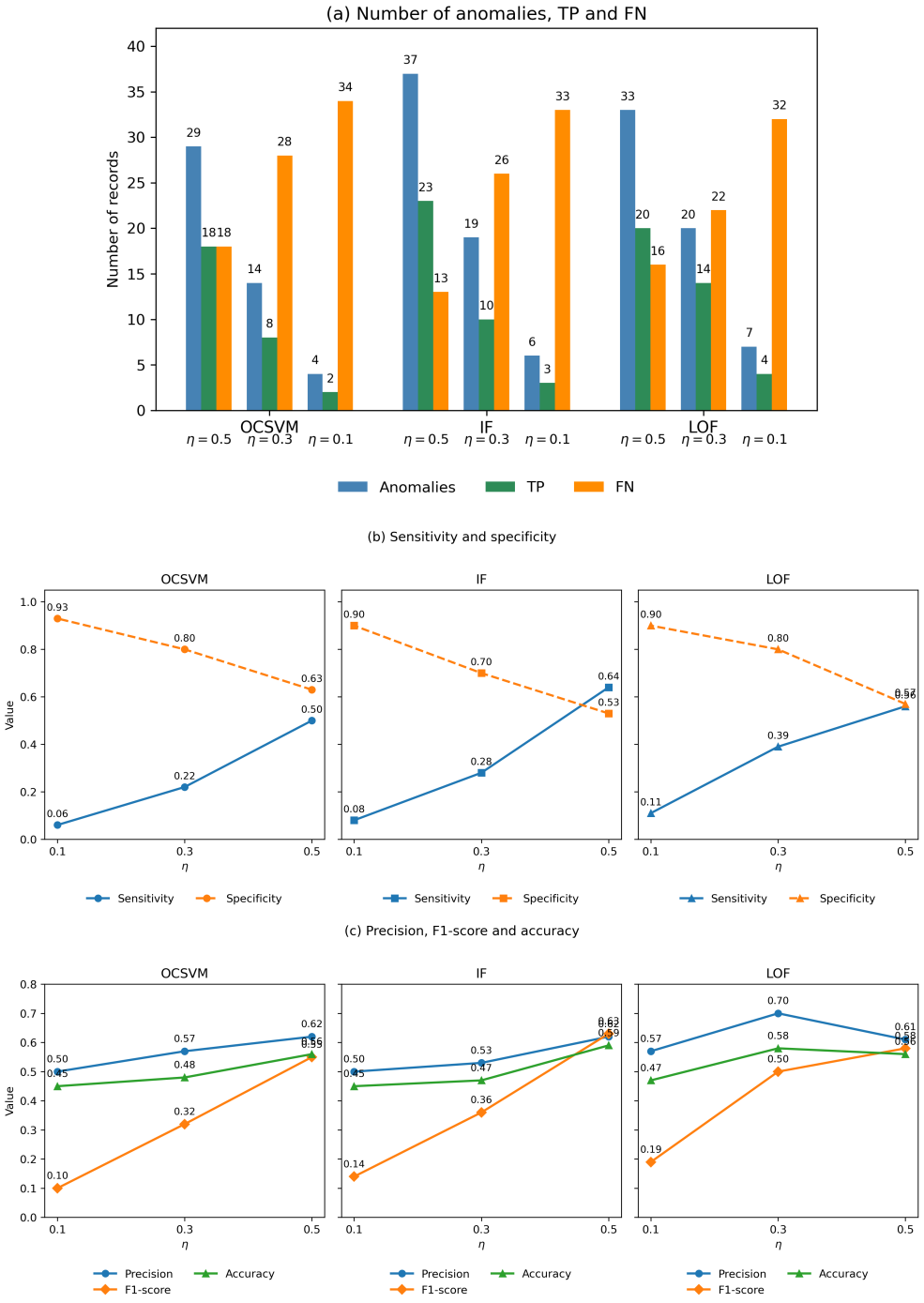


Fig. 11: Performance metrics for anomaly-based methods on data from 2025.

5.2.2 Comparison of alert generation strategies

In this section, we evaluate the three approaches by comparing the number of alerts generated in order to assess their alerting behavior and degree of selectivity. To compare the number of records associated with alerts generated by each method, we used a Venn diagram showing the number of records involved in the alerts triggered by the three approaches.

For the rule-based system, we used a database of logs generated from email notifications sent during the study period. This dataset was then filtered by animal type to retain only alerts associated with records involving NHPs. The resulting dataset comprised 12,869 alerts triggered by 2,715 records, generating a total of 32,362 emails sent.

For the index-based approach, we selected the clusters that generated alerts at a threshold level of $TL = 0.8$, resulting in 47 clusters containing 128 records. For the anomaly-based approach, we selected the anomalous clusters identified by the three anomaly detection methods using a contamination level of $\eta = 0.5$. This resulted in 19 clusters comprising 71 records.

The overlap among the records involved in alerts generated by the three approaches is shown in Figure 12. The rule-based approach involved the largest number of records (2,715), with 2,592 records (95.5%) being unique to this approach. In contrast, only 1 record was unique to the index-based approach, while 18 records were unique to the anomaly-based approach. Forty-three records were identified by all three approaches, indicating a common subset of events consistently flagged regardless of the alerting strategy.

The index-based and anomaly-based approaches shared 50 records, including the 43 records also identified by the rule-based system. The index-based approach shared 120 records with the rule-based approach, whereas the anomaly-based approach shared 46 records with the rule-based system. Overall, the three approaches involved 2,741 distinct records, with the large majority associated exclusively with the rule-based system.

The overlap observed among the three approaches highlights their complementary behavior. While the rule-based system identifies a broad set of records according to predefined surveillance criteria, the index-based and anomaly-based approaches provide more selective mechanisms for highlighting subsets of records based on their attributes and patterns. The fact that each approach also identifies records that are not selected by the others suggests that combining these strategies may increase the diversity of events brought to the attention of surveillance teams. Rather than representing competing alternatives, the three approaches can therefore be viewed as complementary mechanisms that operate at different levels of selectivity and may jointly support a more comprehensive surveillance process.

An additional analysis of the records identified exclusively by the index-based and anomaly-based approaches provided further evidence of their complementary behavior. The records unique to these approaches corresponded to observations of live NHPs with normal behavior, and therefore would not be expected to trigger rules based on reports of sick or dead animals. Nevertheless, the sole record identified exclusively by the index-based approach and 8 of the 18 records identified exclusively by the anomaly-based approach coincided with confirmed YF cases reported in the BMoH database within the 30 day-window following the initial observation. Thus, although these records did not present characteristics that would directly trigger the rule-based system, their inclusion in clusters identified by the index-based or anomaly-based approaches indicates that they occurred in temporal contexts associated with confirmed YF events.

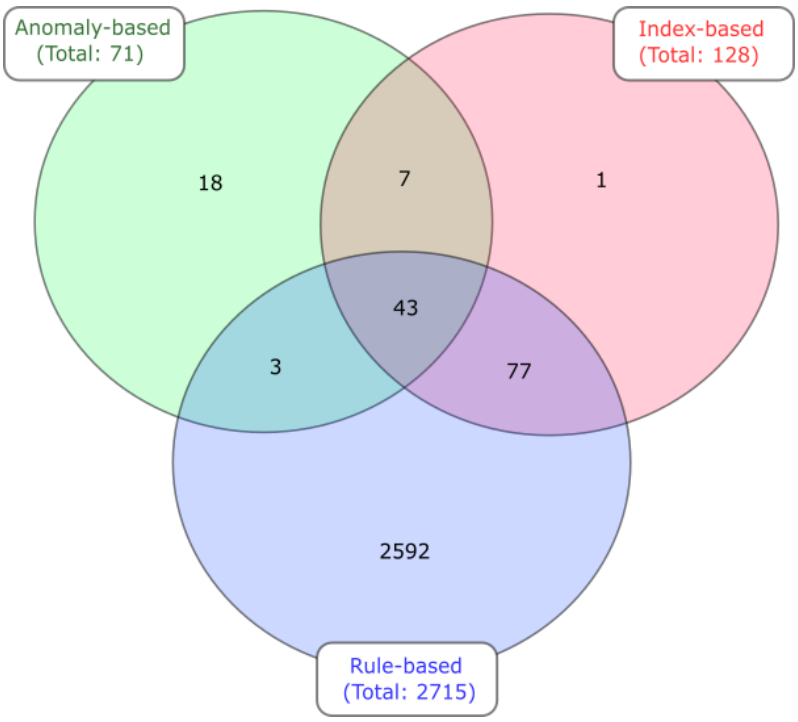


Fig. 12: Venn diagram showing the number of records involved in the alerts triggered by the three approaches.

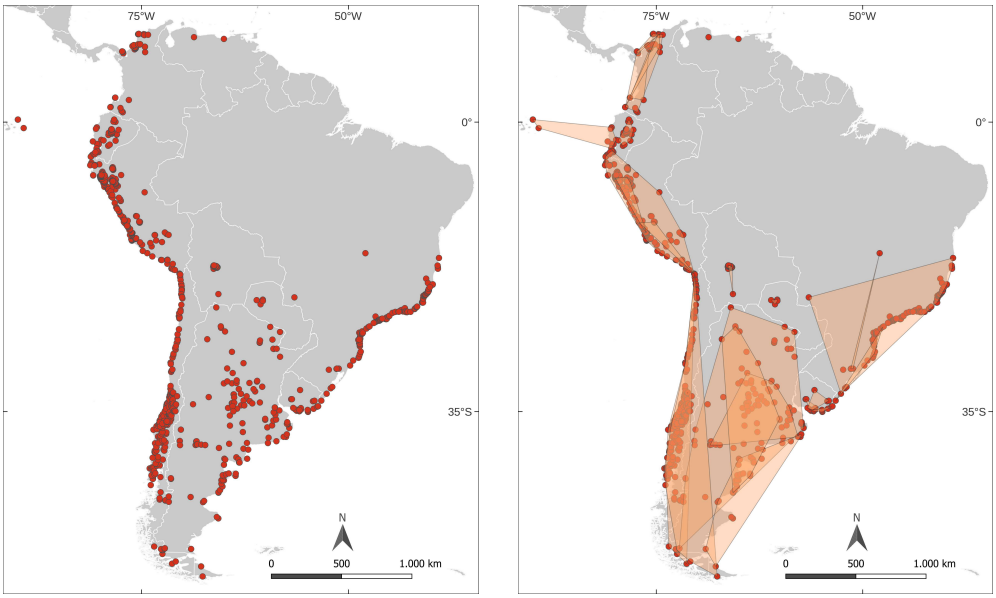
This finding highlights an important distinction between the approaches. The rule-based system primarily evaluates individual records against predefined conditions, whereas the index-based and anomaly-based approaches consider records jointly within spatial and temporal clusters. Consequently, a record involving a live animal with no reported abnormal behavior may still contribute to a cluster whose overall characteristics warrant increased attention. In this sense, the proposed approaches do not replace rule-based alerts but can complement them by identifying patterns that are not necessarily detectable from the characteristics of an individual record alone.

5.3 Case Study 2: Avian Influenza surveillance

In contrast to the YF scenario, the surveillance of avian influenza presents additional challenges due to the limited availability of confirmed cases within the SISS-Geo dataset. During the data export period of the experiment, there was only one confirmed case. As a result, the index-based approach could not be reliably applied, and the analysis relied exclusively on anomaly detection methods to identify atypical patterns in bird-related observations.

In this study [32], we analyzed bird records from the SISS-Geo platform using the anomaly-based approach, and compared the results with official data on AI outbreaks provided by the WOAH [33], between October 2020 and June 2025. Avian influenza is a highly contagious viral zoonosis with a high mutation rate and has the ability to adapt to new

hosts, making early detection essential to prevent its spread to commercial poultry farms and mitigate economic and food security risks [34]. Figures 13a and 13b illustrate the spatial distribution of 1,171 outbreaks (red points) aggregated into 37 epidemiological events (orange polygons), according to WOAH.



(a) Spatial distribution of outbreaks.

(b) Outbreaks grouped into events.

Fig. 13: Spatial distribution of H5N1 outbreaks in South America provided by WOAH.

The spatial and temporal thresholds employed in the clustering analysis were empirically derived from the outbreak grouping structure provided in the WOAH database, also adopted in Brazil. According to the WOAH Terrestrial Animal Health Code [35], an event is defined as a single outbreak or a group of epidemiologically related outbreaks associated with the same disease occurrence, reported from the initial notification through the final report. Therefore, outbreaks sharing the same report identifier (reportID) are considered epidemiologically linked by reporting authorities, although no explicit spatial or temporal criteria are specified by WOAH for this grouping. In this study, these reportIDs were used as reference clusters to infer the typical spatial and temporal extent of epidemiologically related outbreaks.

For each reportID, spatial and temporal distances between neighboring outbreaks were calculated, and the maximum distance observed within each group was extracted. This procedure generated distributions of cluster-specific maximum spatial and temporal distances. Since some reportIDs exhibited exceptionally large maximum distances (1372.36 km), potentially reflecting reporting gaps, under-detection of intermediate outbreaks, or other reporting inconsistencies, the absolute maximum values were considered unsuitable for parameterization. Instead, the median of the cluster-specific maximum distances was adopted as a robust measure of central tendency. The resulting thresholds (303 km and 17 days) were then

used as the maximum spatial and temporal criteria in the clustering procedure. Thus, rather than reproducing an explicit WOAH classification rule, the proposed thresholds represent an empirical approximation of the spatial and temporal scales commonly observed among outbreaks considered epidemiologically related within the WOAH reporting system. Using these parameters, 7,228 SISS-Geo records of birds were grouped into 568 clusters during the clustering phase. These clusters were subsequently characterized and classified as anomalous or non-anomalous using three unsupervised models: One-Class SVM (OCSVM), Isolation Forest (IF), and Local Outlier Factor (LOF).

As illustrated in Figure 14, the anomaly level of each cluster was determined based on the proportion of unsupervised models that identified it as anomalous. Clusters are represented by polygons, whereas their associated records are displayed as points. An anomaly level of 0% corresponds to clusters not identified as anomalous by any model (green polygons and points), 33% corresponds to clusters identified by a single model (yellow polygons and points), and 66% corresponds to clusters identified by two different models (red polygons and points).

A total of 536 clusters were classified as non-anomalous (0%), corresponding to 2,043 records; 25 clusters were classified as anomalous by a single unsupervised model (33%), totaling 82 records; and 7 clusters were classified as anomalous by two unsupervised models (66%), comprising 5,103 records. The results indicate a high concentration of records within anomalous clusters, totaling 5,185 records (71.7%).

A strong spatial association was observed between anomalous clusters and confirmed avian influenza outbreaks. Approximately 92.7% of the outbreaks (152 cases) were located near at least one SISS-Geo record, considering the same spatial and temporal limits used for the clustering stage. It was identified that 77% of the outbreaks (126 cases) were associated with anomalous clusters, while 16% (26 cases) were linked to non-anomalous clusters, and 7% (12 cases) showed no association with SISS-Geo data.

These findings suggest that anomaly detection methods can identify spatiotemporal patterns associated with avian influenza outbreaks, even in scenarios with limited labeled data, since there is a strong correlation between anomalous clusters and outbreaks. The ability to detect clusters that deviate from expected behavior makes this approach particularly suitable for emerging diseases or in underreported conditions.

However, it is important to acknowledge certain limitations in the validation process. The definition of outbreaks provided by WOAH does not fully capture the underlying epidemiological dynamics, which may affect the spatial and temporal alignment between observed clusters and reported events. Additionally, the low frequency of SISS-Geo records and the underreporting of causes of death in bird observations may reduce the sensitivity of the models. Despite these limitations, the results reinforce the potential of anomaly detection as a valuable tool for supporting avian influenza surveillance, particularly in data-scarce environments where traditional, label-dependent methods are not feasible.

5.4 Strengths and Limitations

Each alert generation strategy presents inherent limitations when applied in isolation. Table 4 summarizes the main differences between the alert generation strategies, highlighting their complementary characteristics in terms of interpretability, flexibility, and detection capabilities.

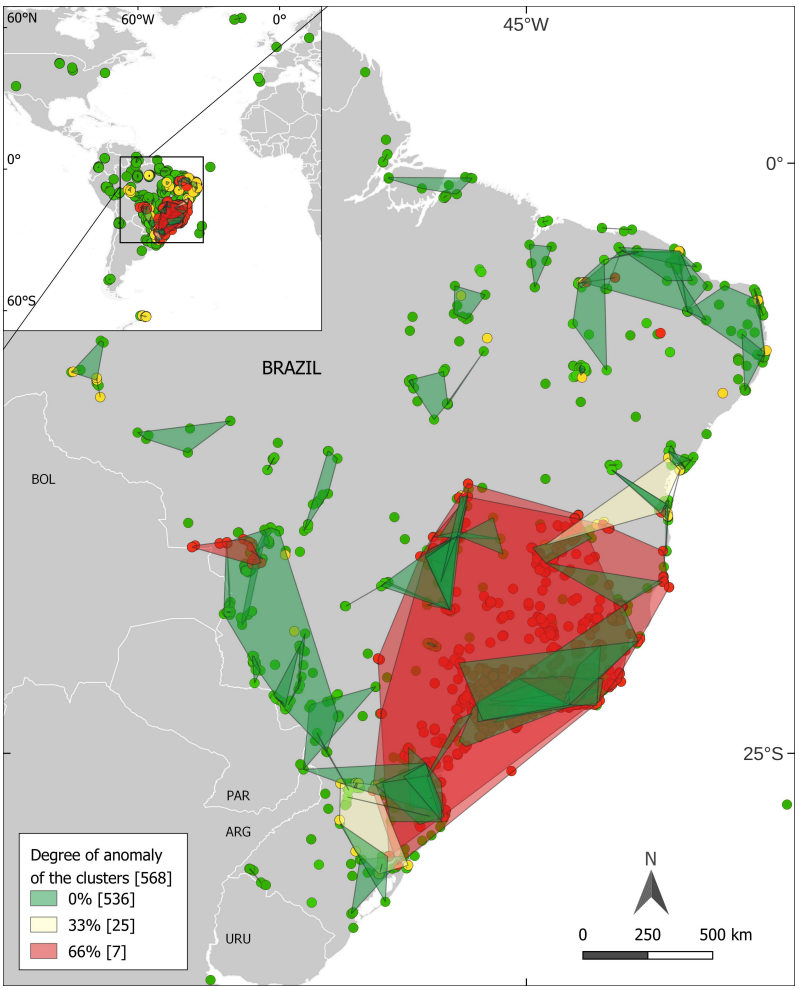


Fig. 14: Association of SISS-Geo records into groups and anomaly classification.

Index-based alerts provide a structured representation of levels of attention by integrating multiple cluster attributes through calibrated weights. However, relying solely on predefined attributes and calibrated weights, considering clusters with a given disease of interest, may not capture all relevant signals. Moreover, the optimization procedure depends on the availability of confirmed disease cases to guide the calibration of attribute weights. When such information is unavailable or incomplete, the optimization process may become less discriminative, reducing the ability of the index to clearly distinguish between critical and non-critical clusters.

On the other hand, anomaly detection methods identify atypical patterns directly from the data without relying on predefined assumptions, providing an additional layer of sensitivity, and helping to detect early warnings. Nevertheless, some detected anomalies may not correspond to epidemiologically relevant events.

Table 4: Comparison between rule-based, index-based, and anomaly-based approaches.

Aspect	Rule-based	Index-based	Anomaly-based
Method type	Fixed rules	Optimization-based	Unsupervised learning-based
Output	Binary alert (alert/no alert)	Continuous alert index (level of attention)	Binary classification (anomalous vs normal)
Interpretability	Highly interpretable and transparent	Explicit contribution of clusters attributes via weights	Model-dependent and less intuitive
Labeled data	Not required	Requires confirmed cases for optimization	Not required
Domain knowledge	Fully defined by expert rules	Objective functions and attributes are domain-driven	Relies primarily on data distribution
Flexibility	Rules must be manually updated	Requires recalibration for new diseases or contexts	Adaptable to different datasets and diseases
Sensitivity to parameters	–	Mainly influenced by optimization setup and threshold level	Strongly affected by contamination rate and model parameters
Detection capability	Effective for predefined events	Effective in identifying known relevant patterns	Effective in detecting atypical or unexpected patterns
Risk of false negatives	May miss events not covered by predefined rules	Increases with higher threshold level	Increases with lower contamination rates
Operational use	Routine surveillance and automatic notification. Supports decision-making	Supports prioritization and decision-making	Supports early warning and exploratory analysis

To highlight the complementary roles of the proposed alert generation strategies, Table 5 summarizes their main characteristics, benefits, and limitations.

The integration of rule-based, index-based, and anomaly-based alert mechanisms offers complementary perspectives for identifying records and clusters of interest, supporting risk assessment and enhancing the robustness, sensitivity, and adaptability of the surveillance system. Although these models operate at the record or cluster level, public health managers often need to interpret their outputs at broader territorial scales, such as municipalities, regions, and operational areas, where surveillance priorities and interventions are defined. By providing complementary information to guide this process, such an integrated approach can strengthen decision-making and support preventive actions before zoonotic pathogens are confirmed, potentially reducing the time between emergence and response.

Table 5: Strengths and limitations of each alert generation strategy.

Alert Strategy	Main Characteristics	Benefits	Limitations
Rule-based	Alerts generated from predefined criteria (e.g., animal deaths and physical symptoms).	Simple interpretation, transparent decision rules, and aligned with established surveillance protocols.	Limited flexibility. It may fail to detect unusual or emerging patterns that fall outside predefined rules.
Index-based	Alerts generated from a composite index combining multiple cluster attributes with calibrated weights.	Combines multiple attributes into a single quantitative measure of cluster criticality, enabling ranking and prioritization of clusters.	Depends on selected variables and calibrated weights, which require confirmed cases whose availability may depend on the time required for diagnosis.
Anomaly-based	Alerts generated by identifying atypical patterns using unsupervised machine learning methods.	Detects unexpected or emerging patterns, it is independent of predefined assumptions, and supports the search for complementary signals.	May generate alerts that are not epidemiologically relevant and is sensitive to data distribution.

6 Conclusion

This study presented the evolution and integration of three complementary alert generation strategies within the SISS-Geo platform for wildlife disease surveillance: rule-based alerts, index-based, and anomaly-based approaches. Together, these methodologies provide complementary alert strategies that operate under different conditions, temporal scales, disease contexts, and levels of selectivity. They complement each other and support improved health surveillance actions.

The rule-based system demonstrated the importance of simple and transparent mechanisms for the rapid communication of known outbreak events. However, the large number of generated notifications and the absence of a systematic prioritization mechanism highlighted the limitations of relying exclusively on fixed criteria. To address these issues, the index-based approach introduced a continuous and interpretable measure of cluster relevance, allowing clusters (group of records) to be ranked and prioritized according to multiple spatial, temporal, and cluster-level attributes. The results obtained in the YF case study demonstrated that the Z-Alert index can effectively support surveillance teams by balancing sensitivity and operational workload through adjustable thresholds.

In parallel, the anomaly-based approach expanded the surveillance capacity by detecting atypical patterns directly from the data without depending on predefined assumptions or labeled datasets. The experiments conducted with YF and avian influenza data showed that unsupervised models are capable of identifying clusters associated with relevant events, particularly in scenarios where confirmed cases are scarce or unavailable. The strong spatial association observed between anomalous bird clusters and official avian influenza outbreaks reinforces the potential of anomaly detection as a proactive surveillance tool for emerging diseases.

The comparative analyses demonstrated that each strategy provides distinct and complementary contributions to wildlife health surveillance. Rule-based alerts offer operational

simplicity and alignment with established protocols; index-based alerts provide structured prioritization and interpretability; and anomaly detection methods contribute exploratory and adaptive capabilities for identifying unexpected events. Their integration therefore establishes a multi-layered surveillance framework that combines expert knowledge, epidemiological evidence, and machine learning techniques to improve the detection and monitoring of potential zoonotic threats.

Despite the promising results, some limitations remain. The effectiveness of the index-based approach depends on the availability of confirmed cases for model's calibration, while anomaly detection methods are sensitive to parameter selection and may generate alerts that are not epidemiologically relevant. Additionally, underreporting and inconsistencies in wildlife surveillance data may affect model performance. Future work includes the incorporation of additional environmental and climatic variables and investigation on models' parameters, to better understand their impact on the model's sensitivity. Expanding validation across different diseases and ecological contexts will also be essential to strengthen the generalizability of the proposed framework.

This work highlights the potential of integrating traditional surveillance strategies with intelligent, data-driven approaches to advance wildlife disease monitoring. By combining rule-based, index-based, and anomaly-based alert mechanisms, the proposed framework provides complementary information for identifying records and clusters of interest, supporting the development of more proactive and scalable surveillance systems. Although the operational benefits of translating model outputs into territorial priorities and management actions remain to be evaluated, these complementary approaches offer a methodological foundation for future decision-support tools aimed at guiding surveillance efforts at municipal, regional, and operational levels. By combining technological innovation with public participation and citizen science, this framework contributes to strengthening wildlife health surveillance within the One Health perspective.

Acknowledge

The authors thank the financial support provided by the Brazilian agencies Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (grants 380656/2024-8, 445493/2023-2 and 311177/2025-5).

References

[1] Jones, K.E., Patel, N.G., Levy, M.A., Storeygard, A., Balk, D., Gittleman, J.L., Daszak, P.: Global trends in emerging infectious diseases **451**(7181), 990–993 (2008) https://doi.org/10.1038/nature06536

[2] Mutairi, S.A.A., Alsuraimi, F.A.H., Ghassap, S.A., Alrawi, N.I., Al-Arej, I.M., Alsubaie, H.F.A., Magliah, R.F., Alshehri, S.A., Nujaifan, N.A.B., Khubrani, W.H., Alfayez, S.F.S., Alanazi, K.M., Al-Shahrani, H.D.T.: The role of surveillance systems in the early detection and response to outbreaks of emerging and reemerging infectious diseases: A comprehensive review. Journal of Angiotherapy **8**(8) (2024) https://doi.org/10.25163/angiotherapy.8810207

[3] Zhang, L., Guo, W., Lv, C.: Modern technologies and solutions to enhance surveillance and response systems for emerging zoonotic diseases. Science in One Health **3**, 100061 (2024) https://doi.org/10.1016/j.soh.2023.100061

[4] Sinclair, J.R.: Importance of a one health approach in advancing global health security and the sustainable development goals. Revue scientifique et technique (International Office of Epizootics) **38**(1), 145–154 (2019) https://doi.org/10.20506/rst.38.1.2949

[5] Deiana, G., Arghittu, A., Dettori, M., Castiglia, P.: One world, one health: Zoonotic diseases, parasitic diseases, and infectious diseases. Healthcare **12**(9) (2024) https://doi.org/10.3390/healthcare12090922

[6] Wakimoto, M.D., Menezes, R.C., Pereira, S.A., Nery, T., Castro-Alves, J., Penetra, S.L.S., Ruckert, A., Labonté, R., Veloso, V.G.: Covid-19 and zoonoses in brazil: Environmental scan of one health preparedness and response. One Health **14**, 100400 (2022) https://doi.org/10.1016/j.onehlt.2022.100400

[7] Gadelha Jr, L.M.R., Siracusa, P.C., Dalcin, E.C., Silva, L.A.E., Augusto, D.A., Krempser, E., Affe, H.M., Costa, R.L., Mondelli, M.L., Meirelles, P.M., Thompson, F., Chame, M., Ziviani, A., Siqueira, M.F.: A survey of biodiversity informatics: Concepts, practices, and challenges. WIREs Data Mining and Knowledge Discovery **11**(1), 1394 (2021) https://doi.org/10.1002/widm.1394 https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/widm.1394

[8] Abdalla, L., Augusto, D.A., Chame, M., Dufek, A.S., Oliveira, L., Krempser, E.: Statistically enriched geospatial datasets of brazilian municipalities for data-driven modeling. Scientific Data **9**(1), 1–12 (2022) https://doi.org/10.1038/s41597-022-01581-2

[9] Heiderich, E., Keller, S., Pewsner, M., Origgi, F.C., Zürcher-Giovannini, S., Borel, S., Marti, I., Scherrer, P., Pisano, S.R.R., Friker, B., Adrian-Kalchhauser, I., Ryser-Degiorgis, M.-P.: Analysis of a european general wildlife health surveillance program: Chances, challenges and recommendations. PLOS ONE **19**(5), 1–27 (2024) https://doi.org/10.1371/journal.pone.0301438

[10] Sharan, M., Vijay, D., Yadav, J.P., Bedi, J.S., Dhaka, P.: Surveillance and response strategies for zoonotic diseases: a comprehensive review. Science in One Health **2**, 100050 (2023) https://doi.org/10.1016/j.soh.2023.100050

[11] Barroso, P., López-Olvera, J.R., Kiluba, T., Gortázar, C.: Overcoming the limitations of wildlife disease monitoring. Research Directions: One Health **2**, 3 (2024) https://doi.org/10.1017/one.2023.16

[12] Beard, R., Wentz, E., Scotch, M.: A systematic review of spatial decision support systems in public health informatics supporting the identification of high risk areas for zoonotic disease outbreaks. International Journal of Health Geographics **17**(1), 1–19 (2018) https://doi.org/10.1186/s12942-018-0157-5

[13] Bird, B.H., Mazet, J.A.K.: Detection of emerging zoonotic pathogens: An integrated one health approach. Annual Review of Animal Biosciences **6**(Volume 6, 2018), 121–139 (2018) https://doi.org/10.1146/annurev-animal-030117-014628

[14] Pan American Health Organization: Health Emergency Information and Risk Assessment. Pan American Health Organization (PAHO/WHO). https://www.paho.org/en/health-emergencies/health-emergency-information-and-risk-assessment Accessed 2026-03-10

[15] Brasil. Ministério da Saúde: Centro de Informações Estratégicas em Vigilância em Saúde (CIEVS). Secretaria de Vigilância em Saúde e Ambiente. https://www.gov.br/saude/pt-br/composicao/svsa/cievs Accessed 2026-03-10

[16] Chame, M., Barbosa, H.J.C., Gadelha, *et. al.*: SISS-Geo: Leveraging citizen science to monitor wildlife health risks in Brazil. Journal of Healthcare Informatics Research **3**(4), 414–440 (2019) https://doi.org/10.1007/s41666-019-00055-2

[17] Chame, M., Abdalla, L., Pinter, A., Romano, A.P.M., Krempser, E., Ramos, D.G., Passos, P.H.d.O., Silva, P.C.L., Silva, G.M.P., Gatti, R.R., Augusto, D.A., Sianto, L.: Primates in SISS-Geo: potential contributions of mobile technology, health surveillance and citizen science to support species conservation in Brazil. Neotropical Primates **26**(2), 80–89 (2020) https://doi.org/10.62015/np.2020.v26.42

[18] Angelo, J.S., Abdalla, L., Augusto, D.A., Chame, M., Krempser, E.: An index-based system for early alerts of potential zoonotic disease outbreaks. Plos One (21) (2026) https://doi.org/10.1371/journal.pone.0356739

[19] Angelo, J.S., Abdalla, L., Augusto, D.A., Chame, M., Krempser, E.: Unsupervised anomaly detection for enhancing wildlife disease surveillance. In: editor (ed.) 2025 12th International Conference on Soft Computing & Machine Intelligence (2025)

[20] Kahariri, S., Thumbi, S.M., Bett, B., Mureithi, M.W., Nyaga, N., Ogendo, A., Muturi, M., Thomas, L.F.: The evolution of kenya's animal health surveillance system and its potential for efficient detection of zoonoses. Frontiers in Veterinary Science **Volume 11 - 2024** (2024) https://doi.org/10.3389/fvets.2024.1379907

[21] TM, O., DS, M., RM, B., AM, C., JM, B., S, M., D., E.: Early warning diagnostics for emerging infectious diseases in developing into late-stage pandemics. Acc Chem Res **54**, 3656–3666 (2021)

[22] Walker, J.W., Han, B.A., Ott, I.M., Drake, J.M.: Transmissibility of emerging viral zoonoses. PLOS ONE **13**(11), 1–12 (2018) https://doi.org/10.1371/journal.pone.0206926

[23] Guo, W., Lv, C., Guo, M., Zhao, Q., Yin, X., Zhang, L.: Innovative applications of artificial intelligence in zoonotic disease management. Science in One Health **2**, 100045 (2023) https://doi.org/10.1016/j.soh.2023.100045

[24] Zhang, Y., Cao, Y., Yang, K., Wang, W., Yang, M., Chai, L., Gu, J., Li, M., Lu, Y., Zhou, H., Zhu, G., Cao, J., Lu, G.: Risk predictive models of healthcare-seeking delay among imported malaria patients in jiangsu province based on machine learning. Zhongguo Xue Xi Chong Bing Fang Zhi Za Zhi **35**(3), 225–235 (2023) https://doi.org/10.16250/j.32.1374.2022290

[25] Chaves, E.C.R., Silva, G.A., Jesus Quintino, A., Neves Trindade Junior, K., Andrade Lima, S.B., Carvalho, C.A.M., Mendonça, M.H.R.: Evaluation of SISS-GEO application as a tool for monitoring reservoir animals of the Yellow Fever Virus (YFV) in Brazil. Electronic Journal Collection Health **12**(9), 3664 (2020) https://doi.org/10.25248/reas.e3664.2020 . (in Portuguese)

[26] Andrade, M.S., Campos, F.S., Oliveira, C.H., Oliveira, R.S., Campos, A.A.S., Almeida, M.A.B., Souza Fonseca, V., Simonini-Teixeira, D., Paixão Sevá, A., Temponi, A.O.D., Magalhães, F.M., Chaves, D.C.C., Pereira, M.A., Lamounier, L.O., Menezes, G.G., Aquino-Teixeira, S.M., Gonçalves-dos-Santos, M.E., Bernal-Valle, S., Müller, N.F.D., Cruz Cardoso, J., Santos, E., Mares-Guia, M.A., Albuquerque, G.R., Romano, A.P.M., Franco, A.C., Ribeiro, B.M., Roehe, P.M., Abreu, F.V.S.: Fast surveillance response reveals the introduction of a new yellow fever virus sub-lineage in 2021, in Minas Gerais, Brazil. Memórias do Instituto Oswaldo Cruz **117**, 220127 (2022) https://doi.org/10.1590/0074-02760220127

[27] Giovanetti, M., Pinotti, F., Zanluca, C., al.: Genomic epidemiology unveils the dynamics and spatial corridor behind the Yellow Fever virus outbreak in Southern Brazil. Science Advances **9**(35) (2023) https://doi.org/10.1126/sciadv.adg9204 https://www.science.org/doi/pdf/10.1126/sciadv.adg9204

[28] Ministério da Saúde. Secretaria de Vigilância em Saúde e Ambiente: Nota Técnica Conjunta Nº 27/2025 - DEDT/DPNI/SVSA. (in Portuguese). https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/notas-tecnicas/2025/nota-tecnica-conjunta-no-27-2025-dedt-dpni-svsa/view

[29] Augusto, D.A., Santos Abdalla, L., Krempser, E., Oliveira Passos, P.H., Garkauskas Ramos, D., Pecego Martins Romano, A., Chame, M.: Fine-grained spatial data-driven ensemble modeling for predicting sylvatic yellow fever environmental suitability in brazil. medRxiv (2026) https://doi.org/10.64898/2026.03.26.26349443 https://www.medrxiv.org/content/early/2026/04/17/2026.03.26.26349443.full.pdf

[30] Ester, M., Kriegel, H.-P., Sander, J., Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: Proc. of the Second Int. Conf. on Knowledge Discovery and Data Mining. KDD'96, pp. 226–231 (1996)

[31] Schubert, E., Sander, J., Ester, M., Kriegel, H.P., Xu, X.: Dbscan revisited, revisited: Why and how you should (still) use dbscan. ACM Trans. Database Syst. **42**(3) (2017) https://doi.org/10.1145/3068335

[32] Spolavori, L.K., Angelo, J.S., Augusto, D.A., Rajão, H., Krempser, E.: Detecção de

anomalias via aprendizagem de máquina no apoio à vigilância de influenza aviária no Brasil. In: E-Vigilância 2025 (2025). (in Portuguese)

[33] Animal Health (WOAH), W.O. https://wahis.woah.org/#/event-management. https://wahis.woah.org/#/event-management

[34] Ministério da Saúde: Influenza Aviária (gripe Aviária). https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/i/influenza-aviaria

[35] World Organisation for Animal Health (OIE): Terrestrial Animal Health Code, 28 edn. World Organisation for Animal Health, Paris, France (2019). https://rr-europe.woah.org/app/uploads/2020/08/oie-terrestrial-code-1_2019_en.pdf