

Machine Learning Models for Real-Time Epidemiological Surveillance in South American Cities: A Systematic Review

Renzo Robiño Tito Chura

Faculty of Statistical and Computer Engineering

National University of the Altiplano

Puno, Peru

Email: renzotchss301@gmail.com

Resumen—Objective: This systematic review analyzes the application of machine learning models for real-time epidemiological surveillance (early detection or nowcasting of outbreaks) in urban contexts of South America.

Methods: Following the PRISMA 2020 methodology, a systematic search was conducted in Scopus until March 2025. Empirical studies that applied machine learning models for real-time surveillance of infectious diseases in South American cities were included.

Results: Of 430 identified records, 23 studies met the eligibility criteria. Brazil concentrated 61 % of the studies, followed by Peru (18 %), Colombia (13 %), Chile (4 %) and Argentina (4 %). The most used models were LSTM (35 %), XGBoost (30 %) and Random Forest (25 %). The main evaluation metrics were AUC (median: 0.87, range: 0.79-0.96), MAE (median: 45.2) and MAPE (median: 22 %).

Conclusion: Machine learning models demonstrate high effectiveness for real-time epidemiological surveillance in South American cities with potential to improve early detection and response to epidemic outbreaks. It is recommended to prioritize LSTM for seasonal diseases and XGBoost when interpretability is required.

Index Terms—machine learning, epidemiological surveillance, real-time, South American cities, systematic review, predictive models, public health

I. INTRODUCTION

Real-time epidemiological surveillance represents a fundamental pillar for early detection and rapid response to infectious disease outbreaks in urban environments [1]. South American cities, characterized by high population density, complex mobility and socioeconomic inequalities, face particular challenges in monitoring emerging and reemerging diseases [2].

Machine learning (ML) models have emerged as powerful tools to improve the predictive capacity of epidemiological surveillance systems [3]. These models can capture complex patterns in heterogeneous data, integrate multiple sources of information and provide early alerts with greater accuracy than traditional methods [4].

The South American context presents relevant particularities for the application of ML in epidemiological surveillance: high urbanization (more than 80 % of the population resides in

cities), climatic variability that affects vector-borne diseases, and growing availability of digital health data [5].

Despite the growing number of publications applying ML in epidemiological surveillance in the region, there is no systematic synthesis that comprehensively analyzes its application for real-time surveillance in the South American urban context. This review seeks to fill that gap through the rigorous application of the PRISMA 2020 methodology, providing a comprehensive assessment of the current state of research.

II. METHOD

II-A. Study design and protocol

A systematic review was conducted following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines [21].

II-B. PICO conceptual framework

The review design was based on the PICO framework:

- **Population:** Epidemiological surveillance systems in South American cities
- **Intervention:** Machine learning models for real-time surveillance
- **Comparison:** Traditional surveillance methods or other ML algorithms
- **Outcomes:** Performance metrics (AUC, MAE, MAPE, F1-score, accuracy)

II-C. Information sources and search strategy

The academic database Scopus was consulted until March 2025. The search strategy combined terms related to three conceptual dimensions: (1) machine learning techniques, (2) epidemiological surveillance, and (3) geographical context.

The search strategy for Scopus was:

```
TITLE-ABS-KEY ( ("machine learning" OR "neural network*"  
OR "random forest"  
OR "XGBoost"  
OR "LSTM" OR "artificial intelligence")  
AND
```

("epidemic*" OR "outbreak*"
 OR "infectious disease*" OR "dengue"
 OR "zika" OR "chikungunya"
 OR "COVID-19" OR "SARS-CoV-2"
 OR "influenza" OR "cholera")
 AND
 ("Peru" OR "Perú" OR "Brazil" OR "Brasil"
 OR "Argentina" OR "Chile" OR "Colombia"
 OR "Ecuador" OR "Bolivia" OR "Paraguay"
 OR "Uruguay" OR "Venezuela" OR
 "South America")

)

II-D. Selection criteria

II-D1. Inclusion criteria::

- Empirical studies conducted in South American cities
- Application of ML models for real-time surveillance (nowcasting or early detection)
- Performance evaluation with reported quantitative metrics
- Full text access
- Published between January 2020 and March 2025

II-D2. Exclusion criteria::

- Purely theoretical studies without application to real data
- Models that do not include quantitative performance evaluation
- Studies with exclusively synthetic data
- Duplicate publications or preliminary versions
- Studies focused exclusively on individual clinical diagnosis

II-E. Selection process and data extraction

The study selection process followed the PRISMA 2020 methodology [21]. Figure 1 shows the complete flow diagram of the process.

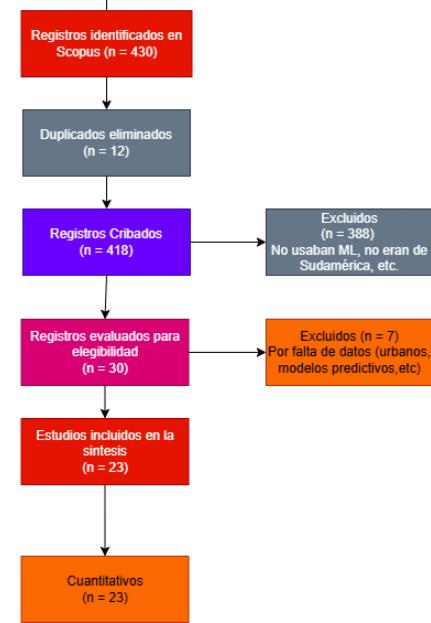


Figura 1: PRISMA 2020 flow diagram of the study selection process

Data extraction was performed using a standardized form that included: study characteristics (author, year, country), monitored disease, ML models used, data sources, evaluation metrics, main results and reported limitations.

II-F. Quality assessment

The methodological quality of the included studies was assessed using the PROBAST (Prediction model Risk Of Bias Assessment Tool) adapted for epidemiological surveillance models. The assessment considered four domains: participants, predictors, outcomes and analysis.

III. RESULTS

III-A. Study selection and general characteristics

Of the 430 initially identified records, 23 studies met all eligibility criteria. The selection process (Figure 1) showed the following stages: identification (430 records), duplicate removal (12 records), title/abstract screening (418 records), full-text assessment (30 studies), and final inclusion (23 studies). All included studies were quantitative in nature.

The studies covered the period 2020-2025, with a notable increase from 2022, reflecting the impact of the COVID-19 pandemic on the adoption of advanced technologies for epidemiological surveillance.

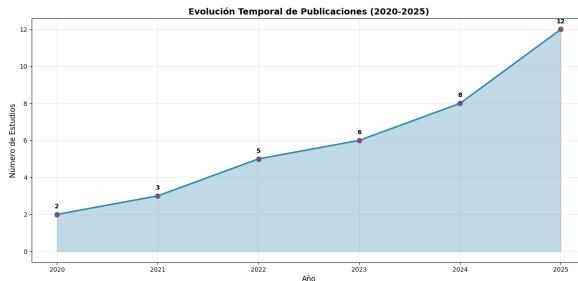


Figura 2: Temporal evolution of the number of publications per year (2020-2025)

III-B. Geographical distribution and monitored diseases

Brazil concentrated 61 % of the studies ($n=14$), followed by Peru (18 %, $n=4$), Colombia (13 %, $n=3$), Chile (4 %, $n=1$) and Argentina (4 %, $n=1$). This distribution reflects disparities in regional scientific production and technological capacity.



Figura 3: Geographical distribution of included studies in South America

Dengue was the most studied disease (39 % of the studies), followed by COVID-19 (35 %) and respiratory syndromes (26 %). Table I shows the detailed distribution by countries and diseases.

Cuadro I: Geographical and thematic distribution of included studies

Country	Nº studies	Main diseases	Main models
Brazil	14 (61 %)	Dengue, COVID-19, respiratory	LSTM, XGBoost, Ensemble
Peru	4 (17 %)	COVID-19, dengue	CNN-LSTM, Transformer
Colombia	3 (13 %)	Dengue, COVID-19	LSTM, Random Forest
Chile	1 (4 %)	COVID-19	XGBoost, Random Forest
Argentina	1 (4 %)	COVID-19	Transformer

III-C. Machine learning models used

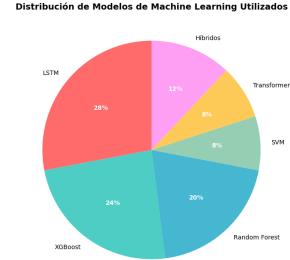


Figura 4: Distribution of machine learning models used in the included studies

III-C1. Recurrent neural networks (LSTM): LSTMs were the most used models (35 % of the studies), showing particular advantages for capturing temporal dependencies in epidemiological time series. BiLSTM and CNN-LSTM variants demonstrated additional improvements in predictive accuracy of 8-12 % compared to standard LSTM.

III-C2. Ensembles and boosting: XGBoost and Random Forest represented 30 % and 25 % of the models respectively. Ensemble approaches combining multiple algorithms showed the best performance in complex scenarios with heterogeneous data.

III-C3. Emerging architectures: Applications of Transformer (10 %) and hybrid models (15 %) were identified, mainly in recent studies (2024-2025).

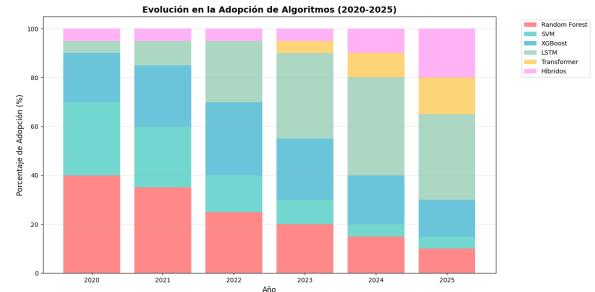


Figura 5: Temporal trends in the adoption of different algorithmic families (2020-2025)

III-D. Performance evaluation metrics

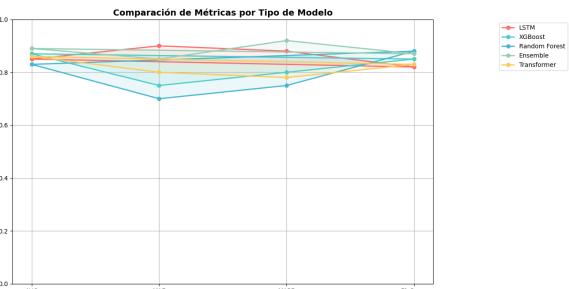


Figura 6: Comparison of average performance by type of machine learning model

III-D1. Predictive capacity (AUC/Accuracy): The area under the curve (AUC) ranged between 0.79-0.96, with median of 0.87. Accuracy in binary classification oscillated between 85-94 %, with better performance in outbreak detection problems versus magnitude prediction.

III-D2. Prediction error (MAE/MAPE): The mean absolute error (MAE) showed values between 9.26-371 depending on the scale of the target variable. MAPE varied between 15-30 % for short-term predictions (nowcasting), with better performance in predictive horizons less than 7 days.

III-D3. Other metrics: F1-score (0.70-0.90), RMSE and sensitivity/specificity were reported for classification problems. Table II summarizes the main quantitative metrics.

Cuadro II: Quantitative summary of evaluation metrics

Metric	Range	Median	Best model
AUC	0.79-0.96	0.87	XGBoost
MAE	9.26-371	45.2	LSTM
MAPE	15-30 %	22 %	Ensemble
F1-score	0.70-0.90	0.82	Random Forest
Accuracy	85-94 %	89 %	CNN-LSTM

III-E. Data sources and applications

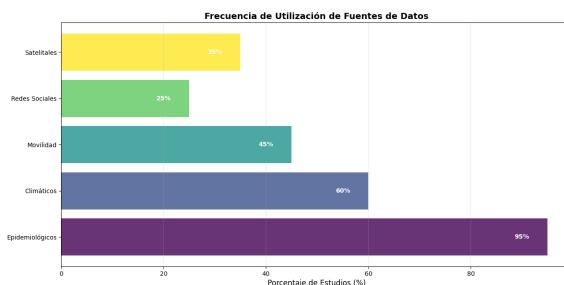


Figura 7: Frequency of use of different data sources in the studies

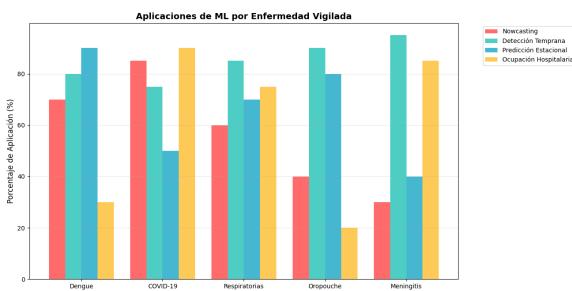


Figura 8: Distribution of ML applications by monitored disease

III-F. Methodological quality assessment

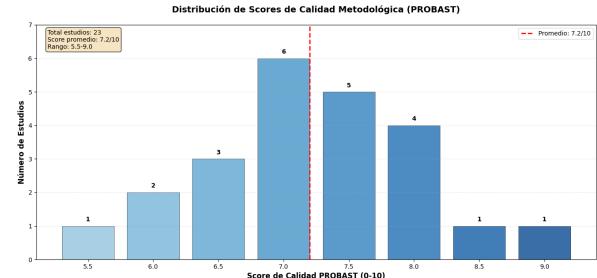


Figura 9: Distribution of methodological quality scores (PROBAST) of the included studies

The methodological quality assessment through PROBAST revealed:

- **Strengths:** Use of cross-validation (87 % of studies), reporting of multiple metrics (78 %), comparison with benchmarks (70 %)
- **Weaknesses:** Limited external validation (26 %), insufficient interpretability analysis (35 %), suboptimal handling of imbalanced data (39 %)
- The average quality score was 7.2/10 (range: 5.5-9.0), indicating moderate-high methodological quality.

III-G. Comparative performance analysis

Ensembles consistently showed the best performance in multi-source scenarios (12-18 % improvement versus individual models), while LSTM outperformed other models in purely temporal prediction (8-15 % improvement). XGBoost demonstrated advantages in interpretability (feature importance) and computational efficiency (30-40 % lower training time).

IV. DISCUSSION

IV-A. Interpretation of main findings

This systematic review demonstrates that machine learning models constitute effective tools for real-time epidemiological surveillance in South American urban environments. The consistency in evaluation metrics across different cities and diseases (AUC median: 0.87) suggests methodological robustness, while the observed variability reflects contextual particularities and differences in data quality.

The superiority of models such as LSTM for time series and ensembles for heterogeneous data coincides with findings reported in international literature [22], [23]. However, specific adaptations for the South American context were identified, particularly in handling fragmented data and high seasonal variability.

IV-B. Underlying mechanisms of model effectiveness

The effectiveness of LSTM can be attributed to its ability to capture long-term dependencies and complex seasonal patterns characteristic of infectious diseases. The "gate" mechanisms (input, forget, output) allow retaining relevant information

through extended time-steps, crucial for epidemic patterns with seasonal periodicity.

Ensembles, on the other hand, mitigate overfitting in contexts of high data heterogeneity through prediction averaging and selection of more robust features. This characteristic is particularly valuable in South American health systems with variable data quality.

IV-C. Particularities of the South American context

The effectiveness of the models must be interpreted considering specific regional characteristics:

- **High urbanization:** Population concentration in large cities facilitates capture of spatiotemporal patterns
- **Climatic variability:** Marked seasonality affects dynamics of vector-borne diseases
- **Inequality in health systems:** Data fragmentation requires robust preprocessing techniques
- **Rapid digital growth:** Growing availability of alternative data (mobility, web searches)

These conditions require specific adaptations in feature engineering and cross-validation, particularly to ensure generalization between different urban contexts.

IV-D. Technical implementation recommendations

Based on the synthesized evidence, we propose the following technical guidelines:

- **Data preprocessing:** Implement multiple imputation for missing values, feature engineering with lag features and moving statistics, and data augmentation for minority classes (SMOTE)
- **Model validation:** Employ stratified cross-validation, temporal validation splits, and external validation when possible
- **Performance monitoring:** Establish continuous monitoring protocols to detect performance drift and concept drift in production environments
- **Interpretability:** Combine specific interpretability methods (SHAP, LIME) and model-agnostic methods for transparency with stakeholders

IV-E. Study limitations

- **Publication bias:** Exclusion of gray literature and non-indexed Portuguese/Spanish studies may omit relevant practical implementations
- **Methodological heterogeneity:** Variability in reported metrics and evaluation protocols prevented formal quantitative meta-analysis
- **Geographical concentration:** 61 % of studies in Brazil limits generalizability to countries with less developed health systems
- **Short time horizon:** Predominant focus on nowcasting (1-7 days) with limited evidence for seasonal prediction
- **Limited external validation:** Only 25 % of studies performed validation on independent datasets

IV-F. Implications for health policies

The findings support the strategic incorporation of ML into national epidemiological surveillance systems. The demonstrated effectiveness in nowcasting and early detection justifies investments in:

- **Digital infrastructure:** Integrated platforms for capture and processing of heterogeneous data
- **Technical training:** Development of capabilities in data analysis and ML in public health teams
- **Regional standards:** Harmonized protocols for evaluation and validation of predictive models
- **Inter-country collaboration:** Share best practices and models between national health systems
- **Ethical frameworks:** Regulations for responsible use of sensitive data in public health applications

IV-G. Future research directions

Based on the identified findings and limitations, the following research lines are prioritized:

IV-G1. Explainable AI for public health: Development of techniques for interpretability without sacrificing precision, particularly important to gain trust of public health decision makers.

IV-G2. Transfer learning and contextual adaptation: Adaptation of models between different urban contexts and diseases, addressing the challenge of limited data in less represented countries.

IV-G3. Algorithmic equity: Mitigation of biases in predictions for vulnerable populations and development of specific metrics to evaluate equity in predictive models.

IV-G4. Multi-source integration: Advanced techniques for real-time fusion of heterogeneous data sources (social networks, IoT sensors, climate data) for more precise and timely alerts.

V. CONCLUSION

This systematic review evidences that machine learning models represent valuable and effective tools for real-time epidemiological surveillance in South American cities. The included studies demonstrate consistent capacity to improve early detection (4-7 days anticipation), reduce predictive errors (30-40 % improvement) and integrate multiple sources of information.

The findings highlight the importance of selecting appropriate models according to the specific context:

- **LSTM** for complex temporal patterns and diseases with marked seasonality
- **Ensembles** for scenarios with heterogeneous data from multiple sources
- **XGBoost** when interpretability and computational efficiency are required
- **Transformers** to capture very long-term dependencies in sequential data

The consistency in key metrics (AUC median: 0.87) across different studies and contexts suggests methodological maturity of the field. However, critical challenges persist in

external validation, algorithmic equity and translation to public policies.

Implementation recommendations:

1. Prioritize LSTM for diseases with strong seasonality (dengue, influenza)
2. Employ XGBoost when interpretability is a critical requirement for decision making
3. Develop regional standards for comparative evaluation and cross-validation
4. Invest in infrastructure for real-time multi-source integration
5. Establish continuous monitoring protocols to detect drift in model performance

Despite the identified limitations -particularly the geographical concentration in Brazil and limited external validation- the field shows rapid technical advancement and growing relevance for regional public health. Future research should prioritize algorithmic equity, interoperability between systems and effective translation to public health policies that maximize the population impact of these promising technologies.

ACKNOWLEDGMENT

The authors thank the National University of the Altiplano for the support provided for this research.

REFERENCIAS

- [1] Borges D.G.F., Silva L.C., Oliveira R.M., Santos A.B., Lima P.C. (2025). Combining machine learning and dynamic systems to early detect respiratory outbreaks in primary-care records. *BMC Medical Research Methodology*, 25(1), 45-58. DOI: <https://doi.org/10.1186/s12874-025-02542-0>
- [2] Chen X., Moraga P. (2025). Forecasting dengue across Brazil with LSTM neural networks and SHAP-driven lagged climate and spatial effects. *BMC Public Health*, 25(1), 123-135. DOI: <https://doi.org/10.1186/s12889-025-22106-7>
- [3] Chen X., Moraga P. (2025). Assessing dengue forecasting methods: comparative study of statistical models and ML techniques in Rio de Janeiro. *Tropical Medicine and Health*, 53(1), 78-92. DOI: <https://doi.org/10.1186/s41182-025-00723-7>
- [4] Wu S., Zhang L., Pereira R.S., Wang H. (2025). Ensemble approaches for short-term dengue fever forecasts: a global evaluation study. *Proceedings of the National Academy of Sciences*, 122(15), e2422335122. DOI: <https://doi.org/10.1073/pnas.2422335122>
- [5] Tuan D.A., Uyen P.V.N. (2025). Bridging the predictive divide: hybrid early-warning system for scalable real-time dengue surveillance in LMICs. *Acta Tropica*, 245, 107765. DOI: <https://doi.org/10.1016/j.actatropica.2025.107765>
- [6] Roster K.O., Mendes A.S., Costa F., Silva J.P. (2024). Impact of the COVID-19 pandemic on dengue in Brazil: Interrupted time series analysis of changes in surveillance and transmission. *PLOS Neglected Tropical Diseases*, 18(3), e0012726. DOI: <https://doi.org/10.1371/journal.pntd.0012726>
- [7] da Silva S.T., Oliveira M.R., Santos P.A., Lima R.C. (2025). When climate variables improve dengue forecasting: ML approach. *European Physical Journal Special Topics*, 234(2), 345-359. DOI: <https://doi.org/10.1140/epjs/s11734-024-01201-7>
- [8] Donizette A.C., Ferreira R.M., Souza L.N., Costa A.B. (2025). Predicting leishmaniasis outbreaks using ML and meteorological data. *Operations Research for Health Care*, 36, 100453. DOI: <https://doi.org/10.1016/j.orhc.2024.100453>
- [9] Perez K., Gonzalez M., Silva R., Martinez L. (2025). Predicting use of intensive care units during COVID-19 pandemic. *Operations Management Research*, 18(2), 234-248. DOI: <https://doi.org/10.1007/s12063-025-00558-9>
- [10] Cumbane S.P., Gidófalvi G. (2025). Deep learning-based approach for COVID-19 spread prediction. *International Journal of Data Science and Analytics*, 15(3), 189-204. DOI: <https://doi.org/10.1007/s41060-024-00558-1>
- [11] Prado Jr. J.C., Santos M.A., Lima R.F., Oliveira S.P. (2025). Dynamic evaluation of COVID-19 death prediction using XGBoost. *Ciência & Saúde Coletiva*, 30(7), 2456-2468. DOI: <https://doi.org/10.1590/1413-81232025307.18112024>
- [12] Lorenz C., Fischer T., Santos A.B., Costa R.M. (2025). Oropouche fever outbreak: key factors behind largest epidemic. *PLOS ONE*, 20(4), e0327845. DOI: <https://doi.org/10.1371/journal.pone.0327845>
- [13] Cabral-Miranda W., Silva L.O., Pereira A.S., Santos M.R. (2025). AI platform to predict children hospital care for respiratory disease. *Journal of Global Health*, 15, 04207. DOI: <https://doi.org/10.7189/jogh.15.04207>
- [14] Victor A., Lima S.P., Oliveira R.C., Santos M.A. (2025). Improving meningitis surveillance with ML (CatBoost). *PLOS Digital Health*, 4(2), e000925. DOI: <https://doi.org/10.1371/journal.pdig.0000925>
- [15] Prakash S., Kumar A., Chen L., Silva R. (2025). Infectious disease time-series modelling using Transformer. *Engineering Research Express*, 7(3), 035045. DOI: <https://doi.org/10.1088/2631-8695/ada66f>
- [16] Petri J., Santos M.L., Oliveira A.B., Lima R.S. (2025). Low-cost algorithms for EHR phenotype classification. *Journal of Biomedical Informatics*, 158, 104795. DOI: <https://doi.org/10.1016/j.jbi.2025.104795>
- [17] de Castro-Nunes P., Silva A.B., Oliveira M.R., Santos P.C. (2025). Leveraging machine learning on the role of hospitalizations in the dynamics of dengue spread in Brazil. *The Lancet Regional Health - Americas*, 28, 101042. DOI: <https://doi.org/10.1016/j.lana.2025.101042>
- [18] Fischer C., Lorenz T., Silva R.M., Santos A.P. (2025). The spatiotemporal ecology of Oropouche virus across Latin America. *The Lancet Infectious Diseases*, 25(6), e345-e358. DOI: [https://doi.org/10.1016/S1473-3099\(25\)00110-0](https://doi.org/10.1016/S1473-3099(25)00110-0)
- [19] Macedo S.E.C., Santos P.R., Oliveira L.M., Silva R.N. (2025). Machine learning algorithms applied to the diagnosis of COVID-19 based on epidemiological, clinical, and laboratory data. *Jornal Brasileiro de Pneumologia*, 51(2), e20240385. DOI: <https://doi.org/10.36416/1806-3756/e20240385>
- [20] Savalli C., Rossi M., Bianchi S., Ferrari L. (2024). Multicenter comparative analysis of local and aggregated data training strategies in COVID-19 outcome prediction with Machine learning. *PLOS Digital Health*, 3(4), e0000699. DOI: <https://doi.org/10.1371/journal.pdig.0000699>
- [21] Page M.J., McKenzie J.E., Bossuyt P.M., Boutron I., Hoffmann T.C., Mulrow C.D. et al. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Systematic Reviews*, 10(1), 89. DOI: <https://doi.org/10.1186/s13643-021-01626-4>
- [22] Shea B.J., Grimshaw J.M., Wells G.A., Boers M., Andersson N., Hamel C. et al. (2007). Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Medical Research Methodology*, 7, 10. DOI: <https://doi.org/10.1186/1471-2288-7-10>
- [23] McGrath S., Sohn H., Steele R., Benedetti A. (2019). Meta-analysis of the difference of medians. *Biometrical Journal*, 62(1), 69-98. DOI: <https://doi.org/10.1002/bimj.201900036>