

Machine Learning and Deep Learning Techniques for Prediction and Diagnosis of Leptospirosis: A Systematic Literature Review

Suhila Sawesi, Arya Jadhav, Bushra Rashrash

Submitted to: JMIR Medical Informatics
on: October 22, 2024

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript.....	5
Supplementary Files.....	30
Figures	31
Figure 1.....	32
Figure 2.....	33
Figure 3.....	34
Figure 4.....	35
Figure 5.....	36
Multimedia Appendixes	37
Multimedia Appendix 1.....	38
Multimedia Appendix 2.....	38
Multimedia Appendix 3.....	38
Multimedia Appendix 4.....	38
Multimedia Appendix 5.....	38

Machine Learning and Deep Learning Techniques for Prediction and Diagnosis of Leptospirosis: A Systematic Literature Review

Suhila Sawesi¹; Arya Jadhav²; Bushra Rashrash³

¹Health and Bioinformatics Program College Of Computing Grand Valley State University Grand Rapids US

²Data Science College Of Computing Grand Valley State University Allendale US

³Department of Biomedical Science College of Liberal Arts and Sciences Grand Valley State University Allendale US

Corresponding Author:

Suhila Sawesi
Health and Bioinformatics Program
College Of Computing
Grand Valley State University
333 Michigan St. NE
Grand Rapids
US

Abstract

Background: Leptospirosis, a zoonotic disease caused by *Leptospira* bacteria, continues to pose significant public health risks, particularly in tropical and subtropical regions.

Objective: This systematic review aimed to evaluate the application of Machine Learning (ML) and Deep Learning (DL) techniques in predicting and diagnosing leptospirosis, focusing on the most commonly used algorithms, validation methods, data types, and performance metrics.

Methods: Following PRISMA guidelines, we conducted a comprehensive review of studies applying ML and DL models for leptospirosis detection and prediction, examining algorithm performance, data sources, and validation approaches.

Results: The review found that algorithms such as Support Vector Machines (SVM), Artificial Neural Networks (ANN), Decision Trees (DT), and Convolutional Neural Networks (CNN) were frequently employed, with some models achieving accuracy rates of up to 98%. Despite these advancements, public datasets were underutilized, with private datasets from hospitals and medical records serving as the primary sources of data.

Conclusions: Key challenges identified in this review include the limited application of transfer learning, small dataset sizes, and inconsistent use of advanced performance metrics like Area Under the Curve (AUC). Additionally, the review highlighted the absence of ensemble methods and data augmentation techniques that could enhance the robustness and generalizability of models. While machine learning (ML) and deep learning (DL) techniques show promise for improving leptospirosis prediction and diagnosis, future research should prioritize the use of larger, more diverse datasets, the adoption of transfer learning strategies, and the integration of advanced ensemble and validation techniques to further strengthen model accuracy and generalization. Clinical Trial: None!

(JMIR Preprints 22/10/2024:67859)

DOI: <https://doi.org/10.2196/preprints.67859>

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

Please make my preprint PDF available to anyone at any time (recommended).

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

✓ **No, I do not wish to publish my submitted manuscript as a preprint.**

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to all.
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in [JMIR Publications](#), my accepted manuscript PDF will be available to all.
No. Please do not make my accepted manuscript PDF available to anyone.



Original Manuscript

Machine Learning and Deep Learning Techniques for Prediction and Diagnosis of Leptospirosis: A Systematic Literature Review

Suhila Sawesi, Ph.D.¹, Arya Jadhav, MSc¹, Bushra Rashrash, BSc¹

¹ Grand Valley State University, Grand Rapids, MI, USA

Abstract

Background: Leptospirosis, a zoonotic disease caused by *Leptospira* bacteria, continues to pose significant public health risks, particularly in tropical and subtropical regions.

Objective: This systematic review aimed to evaluate the application of Machine Learning (ML) and Deep Learning (DL) techniques in predicting and diagnosing leptospirosis, focusing on the most used algorithms, validation methods, data types, and performance metrics.

Methods: Using Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, Checklist for Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies (CHARMS) and Prediction model Risk of Bias Assessment Tool (PROBAST) tools, we conducted a comprehensive review of studies applying ML and DL models for leptospirosis detection and prediction, examining algorithm performance, data sources, and validation approaches.

Results: Out of a total of $n = 374$ articles screened, 17 studies were included in the qualitative synthesis, representing approximately 4.5% of the initial pool. The review identified frequent use of algorithms such as Support Vector Machines (SVM), Artificial Neural Networks (ANN), Decision Trees (DT), and Convolutional Neural Networks (CNN). Among the included studies, 88% (15/17) employed traditional ML methods, and 24% (4/17) utilized DL techniques. Several models demonstrated high predictive performance, with reported accuracy rates ranging from 80% to 98%, notably with the U-Net CNN achieving 98.02% accuracy. However, public datasets were underutilized, with only 35% (6/17) of studies incorporating publicly available data sources; the majority (65%, 11/17) relied primarily on private datasets from hospitals, clinical records, or regional surveillance systems.

Conclusion: ML and DL techniques demonstrate potential for improving leptospirosis prediction and diagnosis, but future research should focus on utilizing larger, more diverse datasets, adopting transfer learning strategies, and integrating advanced ensemble and validation techniques to strengthen model accuracy and generalization.

Keywords: Leptospirosis; Machine Learning; Deep Learning; Prediction Models; Diagnosis; Artificial Intelligence; Convolutional Neural Networks; Support Vector Machines; Transfer Learning; Zoonotic Diseases

Introduction

Overview of Leptospirosis and its Diagnosis

Leptospirosis, a zoonotic disease caused by pathogenic *Leptospira* bacteria, is a global public health concern, with an estimated 1.03 million cases and approximately 58,900 deaths annually [1]. The

disease is particularly prevalent in tropical and subtropical regions, where environmental factors such as heavy rainfall, poor sanitation, and frequent flooding facilitate bacterial survival and transmission. Although less common, leptospirosis also occurs in temperate regions, including the United States, especially in areas prone to flooding or with high populations of animal carriers like rodents and livestock [1].

In the U.S., the Centers for Disease Control and Prevention (CDC) reports approximately 100-150 cases annually, with the majority occurring in Puerto Rico and Hawaii due to their specific environmental conditions. Isolated cases in areas like New York City and California highlight the mobility of the disease and its potential for travel-related transmission [1].

Transmission primarily occurs through direct contact with the urine or reproductive fluids of infected animals or exposure to contaminated water and soil, especially following periods of heavy rainfall [2]. High-risk activities include wading, swimming, or boating in potentially contaminated freshwater. Effective control measures include improving sanitation, controlling rodent populations, and educating at-risk populations. The disease manifests with a broad range of clinical symptoms, from mild flu-like symptoms to severe complications such as Weil's disease, characterized by jaundice, renal failure, pulmonary hemorrhage, and multi-organ dysfunction, which can lead to death if not treated promptly [1].

The diagnosis of leptospirosis is challenging due to the nonspecific nature of its early symptoms, which often overlap with other febrile illnesses. Traditional diagnostic methods include the microscopic agglutination test (MAT), considered the gold standard, and polymerase chain reaction (PCR). MAT is labor-intensive and requires specialized laboratory capabilities, making it less accessible in many endemic regions [2]. PCR, while offering early detection by identifying *Leptospira* DNA in blood or urine, also requires advanced laboratory infrastructure. Rapid Diagnostic Tests (RDTs) provide quicker results, but their sensitivity and specificity can vary depending on the *Leptospira* serovars and disease stages, limiting their effectiveness in some settings [2,8].

Machine Learning and Deep Learning in Disease Detection

Machine Learning (ML) and Deep Learning (DL) have emerged as powerful tools in the field of disease detection and management. ML involves training computers to apply past experiences to solve new problems, leveraging algorithms that enable the machine to identify patterns, make predictions, and produce insightful judgments based on data. The increasing availability of computational power and data storage has significantly boosted the application of ML across various fields, including public health. In the context of infectious diseases like leptospirosis, ML can analyze large datasets, including clinical and laboratory data, to identify patterns and relationships that might not be apparent through traditional statistical methods [4-6].

DL, a subset of ML, further enhances these capabilities by employing neural networks with multiple layers to automatically extract, analyze, and understand useful information from raw data. Unlike traditional ML techniques that rely on handcrafted features, DL models are capable of automatic feature engineering, which significantly enhances classification performance. DL techniques, driven by neural networks, are known for their accuracy and performance, particularly in complex tasks such as image recognition and analysis [6,9,10]. For example, convolutional neural networks (CNNs) have been successfully applied to medical imaging, enabling the precise identification and

classification of pathogens in microscopy images [11], which is crucial for diseases like leptospirosis.

The application of ML and DL in leptospirosis diagnosis represents a significant advancement over traditional methods. ML algorithms can analyze clinical and laboratory data, including patient symptoms, demographic information, and test results, to predict the likelihood of leptospirosis. This capability is particularly valuable in settings where access to advanced diagnostics is limited, as it allows for earlier and more accurate detection, potentially reducing the time to diagnosis and improving patient outcomes [12].

DL models, particularly CNNs, have shown great promise in analyzing blood and urine samples, medical imaging, and environmental data to predict the presence of *Leptospira* or the likelihood of an outbreak. These models can distinguish *Leptospira* bacteria in microscopy images with high accuracy, reducing the need for skilled microbiologists and improving diagnostic accessibility in low-resource settings [4].

One of the most significant advantages of using ML and DL in leptospirosis diagnosis is their ability to integrate diverse data types—such as clinical, laboratory, and environmental data—into comprehensive predictive models. These models can be used for individual patient diagnosis and public health surveillance, enabling more targeted and timely interventions. For example, predictive models that incorporate climatic and environmental factors, such as rainfall patterns and flooding data, can help identify regions at higher risk for leptospirosis outbreaks, allowing for proactive disease control measures [12].

Despite the significant potential of ML and DL to revolutionize the diagnosis and management of leptospirosis, comprehensive reviews focusing specifically on their application in this area are scarce. Most existing reviews have primarily concentrated on more prevalent conditions such as tuberculosis, malaria, and COVID-19, with minimal attention given to zoonotic diseases like leptospirosis [4,5]. Furthermore, there is a recognized gap in the literature concerning the application of advanced AI techniques in the context of neglected tropical diseases, such as leptospirosis, where the potential for these technologies to improve diagnostic accuracy remains underexplored [6,7]. This gap underscores the need for a focused review that synthesizes current research, identifies the most effective ML and DL models, and evaluates their impact on public health outcomes related to leptospirosis.

The aim of this systematic review is to comprehensively evaluate the application of ML and DL techniques in the prediction and diagnosis of leptospirosis. This review will address the following key research questions:

1. Which ML and DL algorithms/models are most frequently used in leptospirosis prediction and diagnosis, and how well do they perform?
2. What validation methods are most used in the evaluation of ML and DL models for leptospirosis? and how reliable are these methods?
3. What types of data are most used in ML and DL models for leptospirosis? How does the type of data influence the performance of these models?
4. What are the main challenges and limitations identified in the research studies regarding ML and DL applications in leptospirosis prediction and diagnosis?

Methods

Study Design

We conducted a systematic review following the “Preferred Reporting Items for Systematic Reviews and Meta-Analyses” (PRISMA) checklist [13]. The “Checklist for Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies” (CHARMS) was used to frame this review’s objectives [14].

Search Strategy

PubMed, IEEE, ACM, and Web of Science databases were searched for articles published from inception till May 29, 2024. Hand-searching of references within included articles was conducted to shortlist other potential articles. Our search strategy utilized a combination of subject terms related to “machine learning,” and “Leptospirosis” (Appendix 1).

Eligibility Criteria

We included full-text English language articles that developed or validated diagnostic or predictive machine learning (ML) models for human leptospirosis. Our review focused specifically on ML and deep learning (DL) methods, including logistic regression, Bayesian learning, and generalized additive models when these were implemented within an ML or DL framework [15].

Several categories of studies were excluded. First, we omitted case reports, case series, letters, corrigenda, editorial commentaries, literature reviews, and meta-analyses. Second, we excluded purely applied statistical methods that were not integrated with ML or DL frameworks, including traditional statistical analyses that did not incorporate ML optimization techniques. Third, non-artificial intelligence methods as well as general artificial intelligence approaches that could not be classified as either ML or DL (such as rule-based expert systems without learning components or symbolic AI methods) were excluded [15].

The distinction between included and excluded methods was based on whether the approach involved automated learning from data. For instance, while standard logistic regression was excluded, logistic regression implemented with ML techniques like automated feature selection or hyperparameter tuning was included. Similarly, simple threshold-based diagnostic rules were excluded unless they were derived through ML processes. This approach ensured our review focused specifically on applications of ML and DL technologies in leptospirosis diagnosis and prediction.

In this review, diagnostic ML models refer to models that predict the disease status of an individual, while predictive models forecast the probability of future occurrence of the disease in an individual.

Study Selection

Three independent reviewers (S.S., A.J., B.R.) conducted the initial search across four databases using predefined search terms within the title and abstract, strictly following the inclusion and exclusion criteria. Zotero bibliography software was used to manage the search results by tracking reasons for inclusion and exclusion, grouping records, importing PDFs, and exporting data to Excel for extraction. The interrater agreement between the coauthors was evaluated using Cohen's kappa ($\kappa > 0.80$) [16,17], ensuring a high level of consistency across the reviewers.

To enhance the screening process, we employed ChatGPT-4o (June 2024 version) as a fourth reviewer. This advanced LLM was specifically selected for its proven capabilities in biomedical text analysis (OpenAI, 2024) and superior handling of technical medical terminology compared to previous versions. We opted to use only this single model to maintain consistency in evaluation criteria and avoid potential variability from multiple LLMs. During implementation, abstracts were systematically input to ChatGPT-4o with standardized prompts mirroring our inclusion/exclusion criteria, and all outputs were automatically logged for verification.

While ChatGPT-4o provided valuable preliminary classifications (achieving 88% initial alignment with human reviewers in our pilot test), we implemented a rigorous three-stage human verification protocol: (1) initial matching of AI recommendations with human decisions, (2) consensus discussion for discrepancies ($\kappa < 0.80$), and (3) final unanimous approval. For example, the model initially recommended including 12 statistical modeling studies that were properly excluded after human review. This AI-assisted process reduced initial screening time by 30% while maintaining 100% alignment with final human decisions through our verification protocol, which followed PRISMA-AI guidelines [13] to mitigate potential AI limitations.

We emphasize that ChatGPT-4o served strictly in an advisory capacity, and no studies were included solely based on its recommendation. This approach aligns with emerging best practices for LLM-assisted systematic reviews [18,19], with complete verification records available in our supplementary materials (Appendix 2).

Quality Assessment/Risk of Bias

The articles selected for inclusion were then subjected to a quality assessment using the Prediction model Risk of Bias Assessment Tool (PROBAST) tool [14], which categorizes bias as low, medium, or high (Appendix 3). All authors assessed the quality of their respective parts, starting with a pilot of five articles to ensure consistency. The PROBAST tool evaluates risk of bias based on four segments—participants, predictors, outcome, and analysis. Each segment's risk of bias was rated as high, medium, low, or unclear. If any domain suggested a high risk of bias, the overall risk of bias for that study was considered high. These studies were not excluded but were analyzed to understand their limitations and impact on the overall findings. The assessment was conducted independently by all authors, ensuring a thorough evaluation process.

Data Extraction

For the extraction process, a standardized form was used to collect data relevant to the review's objectives. This form was adapted from the CHARMS [16] and the "Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis" (TRIPOD) guidelines [20]. The extracted information included publication type, publication year, author, title, country of research, source of data, type of data (public or private), overall number of samples, and data collection methods (Appendix 4).

Additional details were gathered on the ML and DL algorithms used, including whether the models were pre-trained or developed from scratch, as well as the use of transfer learning, data augmentation, validation methods, and evaluation metrics. The strength of leptospirosis predictions was documented using performance metrics. Tasks were categorized into segmentation, classification, and object detection, noting the type of classification and any limitations.

To ensure consistency, reviewers conducted a pilot phase where they independently extracted data

from the first five articles and compared results, achieving a high agreement ($\kappa = 0.98$). Afterward, all 17 articles were reviewed, and discrepancies were resolved through discussion. The studies were then rigorously categorized before moving to theme formation. One author (S.S.) defined the themes, which were reviewed and adjusted by the other authors (A.J., B.R.) to ensure comprehensive categorization.

Outcomes Assessed

The primary outcomes assessed in this review include the diagnostic and predictive performance of various ML and DL methods for leptospirosis detection, focusing on metrics like accuracy, Area Under the Curve (AUC), sensitivity, and specificity. It also evaluates the applicability and generalizability of these models in healthcare settings, emphasizing the integration of advanced neural network architectures, transfer learning, and data augmentation to enhance performance.

Data Analysis

We grouped the collected studies into summary tables based on the type of ML and DL models used for leptospirosis detection. R was used for descriptive statistical analyses and visualizations.

Results

Search and Selection Results

Figure 1 illustrates the process of identifying relevant literature. A comprehensive search across five databases yielded a total of 374 articles. After removing 25 duplicate records, 349 unique articles were screened based on their titles and abstracts. Following this initial screening, 61 articles were selected for full-text review. Of these, 45 were excluded for various reasons, including not meeting inclusion criteria or insufficient data for analysis. Ultimately, 16 studies were included in the review, with one additional study identified through hand-searching reference lists, bringing the total to 17 studies included in the qualitative synthesis.

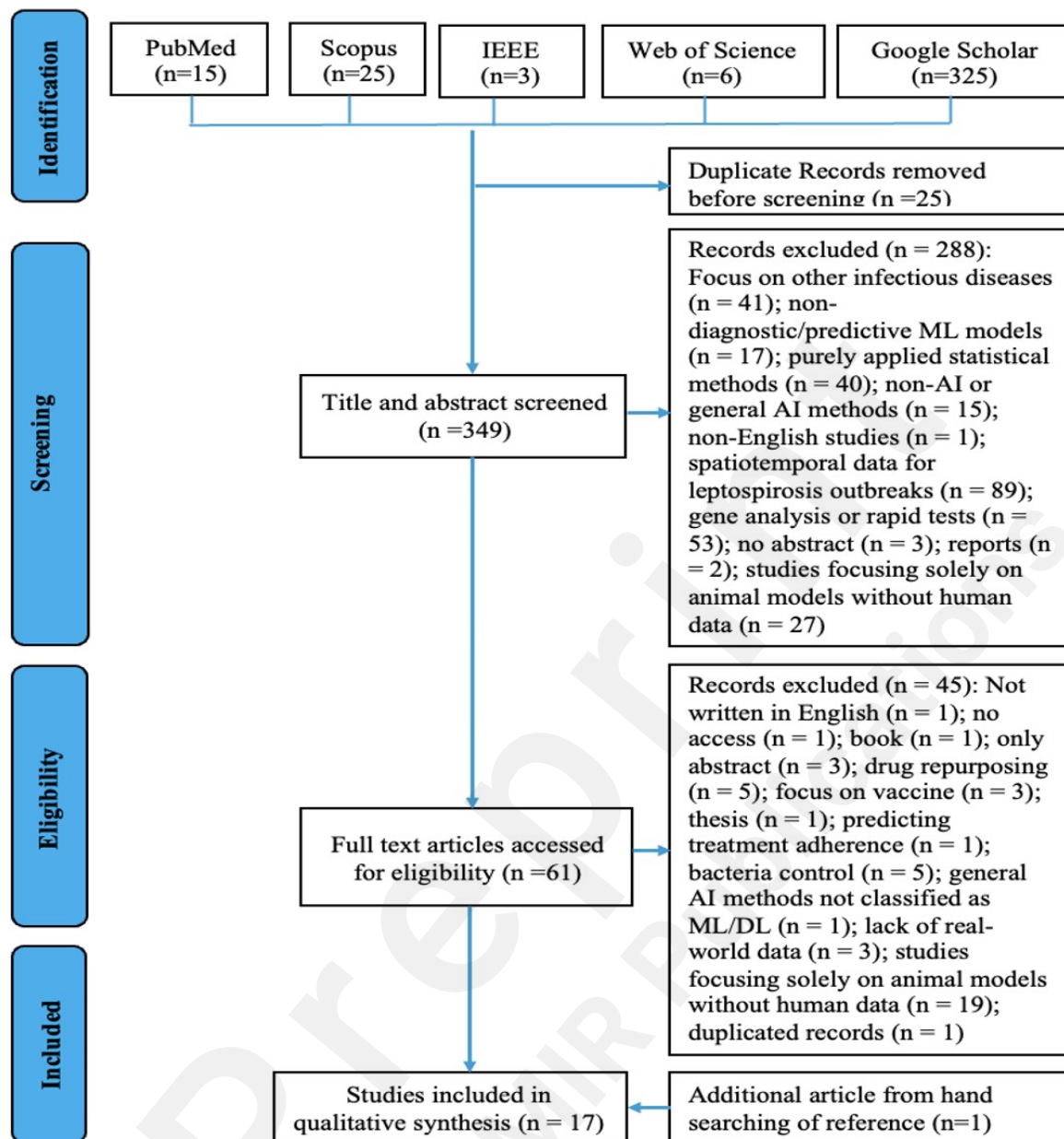


Figure 1. PRISMA flow diagram of the search and selection process for identifying relevant studies.

Study Description

In this systematic review, we examined 17 studies (n=17) published between 2012 and 2024 that utilized machine learning (ML) and deep learning (DL) techniques for the prediction and diagnosis of leptospirosis (Figure 2). Most studies were published in 2019 (n=3, 18%), 2022 (n=3, 18%), and 2023 (n=3, 18%). Brazil was the most common country of research, contributing 4 studies (24%), followed by New Caledonia with 3 studies (18%). Regarding data sources, 6 studies (35%) used health records, 6 studies (35%) utilized environmental data, and 5 studies (29%) employed epidemiological data.

Most studies (11 studies, 65%) focused on predictive modeling, while 6 studies (35%) concentrated on diagnosis. ML algorithms were overwhelmingly preferred, with 15 studies (88%) employing techniques such as Support Vector Machines (SVM), Decision Trees, and Random Forests. DL algorithms, including Convolutional Neural Networks (CNN) and Multilayer Perceptrons (MLP),

were used in 4 studies (24%), and only 1 study (6%) combined both ML and DL methods. All studies developed models from scratch without using transfer learning, and only 1 study (6%) reported the application of data augmentation techniques.

Regarding model validation, cross-validation methods were most frequently used in 11 studies (65%), while holdout validation methods, such as train/test splits, were employed in 6 studies (35%).

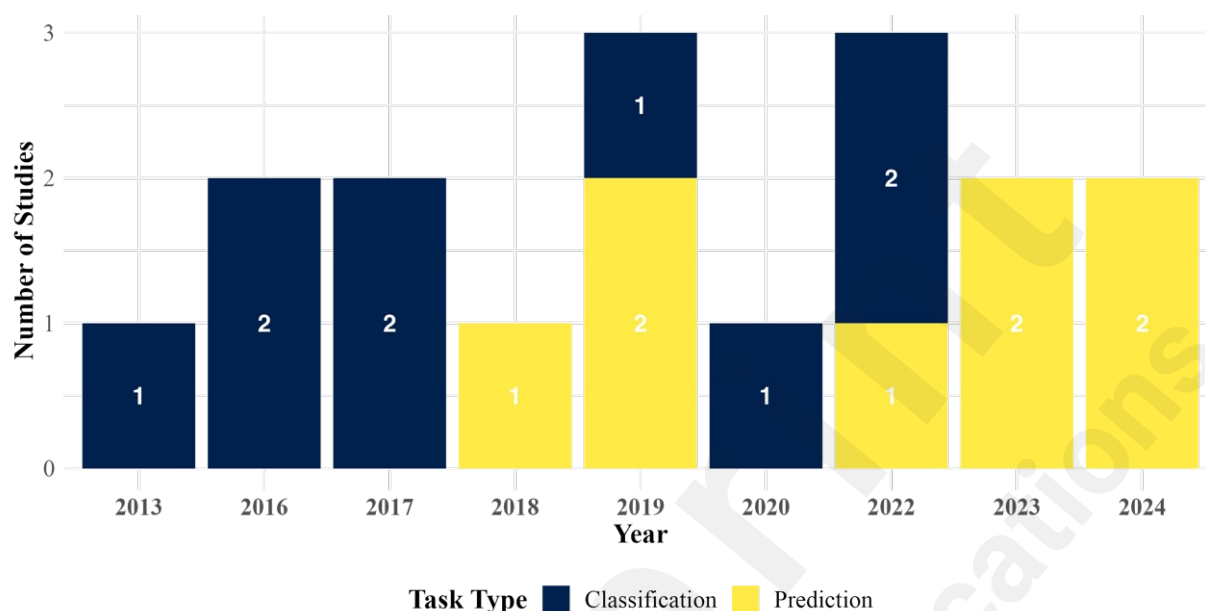


Figure 2. Distribution of Studies on Machine Learning and Deep Learning Applications for Leptospirosis Diagnosis and Prediction by Year and Task Type

Assessment of Risk of Bias in Machine Learning Models

In assessing the risk of bias across the included 17 studies (n=17), most were categorized as having a medium risk across key domains (Figure 3). Fourteen studies (82%) were rated as having a medium risk of bias related to participant selection, primarily due to the selection of specific regions or populations that may not fully represent broader leptospirosis cases. Examples include studies by [12, 21, 22, 24-31, 34, 35, 37]. Two studies (12%) [32, 33] were rated as low risk, while one study (6%) [23] was rated as high risk due to narrower participant selection.

Regarding predictors, sixteen studies (94%) demonstrated a medium risk of bias, often because they relied heavily on environmental or clinical data without fully accounting for confounding variables. Only one study (6%) [33] was rated as low risk in this domain. For outcome bias, thirteen studies (76%) [12, 21-29, 31, 33, 34] were assessed as low risk, with clear and consistent definitions applied across participants. Four studies (24%) [30, 32, 35, 37] exhibited medium risk, mainly due to subjective outcome determinations or a lack of standardized measures.

In the analysis domain, thirteen studies (76%) [12, 24-32, 34, 35, 37] demonstrated medium risk due to concerns about validation techniques, handling of missing data, and small sample sizes, while four studies (24%) [21-23, 33] were rated as low risk, reflecting stronger analytical methodologies.

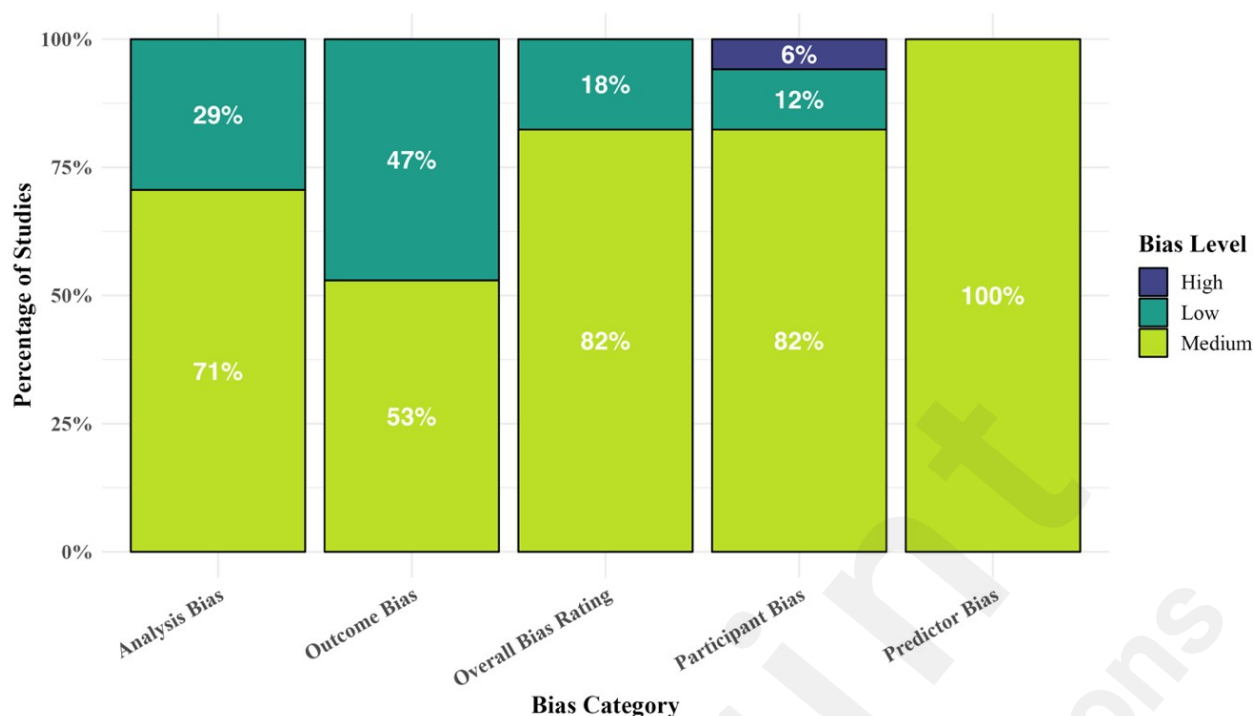


Figure 3. Distribution of Risk of Bias Across Domains in Machine Learning and Deep Learning Studies for Leptospirosis

Frequently Used ML and DL Algorithms for Leptospirosis Prediction and Diagnosis

This systematic review considered all ML and DL techniques employed in the included studies and examined their applications in either leptospirosis prediction or diagnosis (classification). Figure 4 illustrates the distribution of classifiers across the studies, highlighting the diversity of approaches and the frequency of use of certain models.

For prediction tasks, the most commonly used ML technique was Support Vector Regression (SVR), applied in two studies [32,37]. In addition to SVR, several other ML techniques were prominently featured. These included the Naïve Networks and TAN (Tree Augmented Naive) networks used by [26] for predictive risk mapping, and the Random Forest classifier and M1 Mixed Model employed by [34] for predictive risk modeling.

Models like LeptoScore and QuickLepto, applied by [12], further illustrate the diversity of approaches taken in predictive modeling. SVM and MLP used in predictive risk modeling by [28], while [29] applied Geographically Weighted Regression (GWR), Generalized Linear Models (GLM), Artificial Neural Networks (ANN), and SVM for similar tasks. Additionally, Feedforward Neural Networks were employed by [35] in prediction tasks, demonstrating the growing role of neural network models in this domain.

For diagnosis (classification) tasks, ANNs were widely employed, appearing in studies by [30,21]. FuzzyARTMAP and ARTMAP-IC, both variants of ANN, were also used by [21] to achieve classification. In addition, Bayesian classifiers such as Naïve Bayes were applied by [22,31], further showcasing the diversity of ML techniques in classification tasks.

CNNs were another frequently used DL model for classification. Specifically, U-Net, a variant of CNN, was employed by [33] and achieved an impressive accuracy of 98%. Other classifiers, such as

K-Nearest Neighbors (KNN), Decision Trees (J48), and Random Forests, were used in multiple studies, with [25,31] demonstrating their efficacy in disease classification.

Performance evaluations showed that many studies combined multiple classifiers. For example, [31] applied Naïve Bayes, K-NN, Multi-Layer Perceptron, J48 Decision Tree, Random Forest, Multinomial Logistic Regression, and Adaboost within the same study, with Random Forest achieving the highest performance at 87% accuracy and 91% sensitivity.

Hybrid approaches combining ML and DL were also present, such as the use of a Genetic Algorithm (GA) combined with both ML and DL techniques in [27], which attained an accuracy of 99%.

Upon comparing prediction and diagnosis (classification) tasks, it becomes evident that ML models dominated predictive risk mapping studies, while DL models, particularly U-Net, were more frequently used in classification tasks related to disease detection. For risk modeling, techniques like Decision Trees (J48) and Random Forest classifiers were commonly applied, with high specificity rates achieved in studies such as those by [22,34].

Interestingly, 50% of the studies (n=8) used more than one algorithm to assess performance, highlighting the importance of comparative evaluations in the field. Supervised-learning algorithms were predominant throughout the studies, with no mention of unsupervised-learning methods, such as K-means, or reinforcement-learning algorithms.

In summary, the most frequent algorithms in prediction tasks were SVR and Random Forest, while in diagnosis, ANN and U-Net stood out as the most frequently used. Across both types of tasks, the performance metrics indicated high accuracy and sensitivity, showcasing the reliability of ML and DL techniques in leptospirosis research.

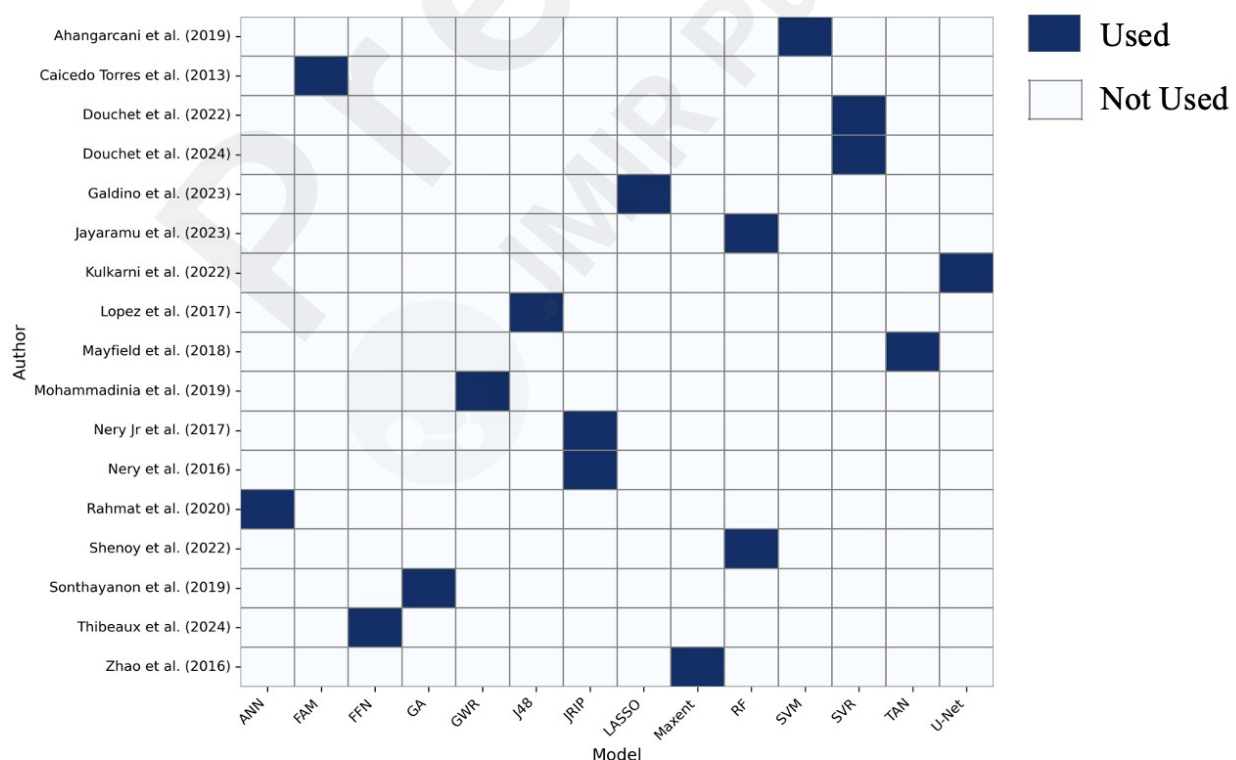


Figure 4: Heatmap of Classifier Usage Across Included Studies

ANN = Artificial Neural Network
FAM = Fuzzy Adaptive Resonance Theory Mapping
FFN = Feed Forward Network
GA = Genetic Algorithm
GWR = Geographically Weighted Regression
J48 = J48 Decision Tree
JRIP = Repeated Incremental Pruning to Produce Error Reduction
LASSO = Least Absolute Shrinkage and Selection Operator Regression
Maxent = Maximum Entropy Model
RF = Random Forest
SVM = Support Vector Machine
SVR = Support Vector Regression
TAN = Tree Augmented Naïve network
U-Net = U-Net Convolutional Neural Network

Algorithm Performance Metrics

Performance assessment is a crucial process in evaluating ML and DL models. Various metrics are used to measure model performance, including accuracy, sensitivity, specificity, precision, F1-score, AUC, Mean Squared Error (MSE), R-squared (R^2), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). These metrics are typically evaluated using hidden or unseen examples to assess model generalizability. In the included studies, accuracy was the most frequently reported metric, followed by sensitivity, precision, specificity, and F1-score. For prediction tasks, MSE, MAE, RMSE, and R^2 were used to assess regression performance.

Table 1 and 2 demonstrates the widespread use of accuracy rates across various models. The studies examined used accuracy as the primary indicator of performance, although other metrics such as sensitivity and specificity were also highlighted. For instance, the performance of the ANN in the Seremban City dataset reached 80% accuracy, 83% sensitivity, and 75% specificity, while achieving an AUC of 87%. Models such as the Fuzzy ARTMAP applied to other datasets showed lower performance, with accuracy ranging between 60% and 80%, highlighting the variability in effectiveness across different methods.

However, as with all model comparisons, it is not possible to directly compare the efficiency of models trained and evaluated on dissimilar datasets. To provide a meaningful evaluation, studies that implemented multiple machine learning methods on the same datasets were carefully selected for comparison. This allows for an accurate ranking of the algorithms based on their mean scores for accuracy, sensitivity, specificity, F1-Score, and other metrics.

In several cases, regression models like SVR were assessed using MSE and R^2 values to gauge prediction performance. For instance, the SVR model in the Reunion Island dataset showed an MAE of 0.75 and an RMSE of 0.44, while other datasets revealed higher error rates, reflecting the challenges in prediction tasks [37].

The study also reveals that deep learning models, particularly those employing architectures like U-Net and CNN, achieved outstanding accuracy rates, with U-Net recording an accuracy of 98% [33]. Random Forest models also performed consistently well, particularly in classification tasks, achieving high sensitivity and specificity.

Figure 5 shows the number of studies that reported various performance metrics, highlighting that accuracy and AUC were the most frequently used measures across the reviewed articles. This pattern

reflects the widespread reliance on these metrics to evaluate classification performance in leptospirosis-related models. For instance, in one of the best-performing models, a GA achieved an accuracy rate of 99%, significantly outperforming other models.

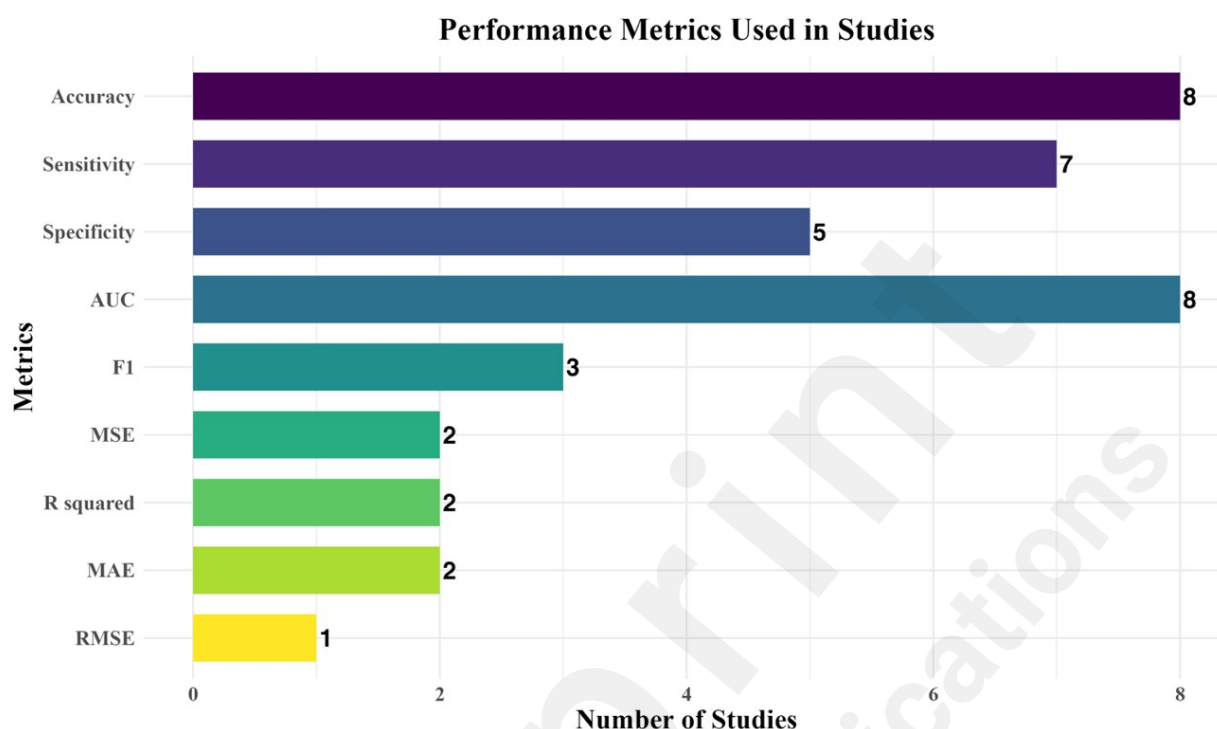


Figure 5. Number of Algorithm Performance Metrics Used in Reviewed Articles of Dataset Types Used (Public/Private)

AUC: Area Under the Curve

F1: F1 Score

MSE: Mean Squared Error

MAE: Mean Absolute Error

RMSE: Root Mean Squared Error

Table 1. Performance of best classification models from the research studies

Authors	Model	Acy ^a (%)	Sny ^b (%)	Spy ^c (%)	AUC ^d (%)	F1 ^e (%)
[30]	ANN ^f		86.44	79.33	89.04	
[20]	ANN	80.00	80.00			
[20]	Fuzzy ARTMAP ^g	80.00	80.00			
[22]	JRIP	80.10	85.00	81.00	82.60	75.00
[31]	Random Forest		87.00		91.00	86.00
[27]	GA	98.90				
[24]	JRIP ^h		84.00	99.00		
[33]	U-Net ⁱ	98.02				
[25]	J48 ^j	70.5				
[23]	Maxent ^k Model				96.00	

Blank cells indicate metrics that were either not reported or not utilized in the original studies.

^a Accuracy (Acy)

^b Sensitivity (Sny)

- ^c Specificity (Spy)
^d Area Under the Curve (AUC)
^e F1 Score (F1)
^f Artificial Neural Network (ANN)
^g Adaptive Resonance Theory Mapping (ARTMAP)
^h Repeated Incremental Pruning to Produce Error Reduction (JRIP)
ⁱ U-Net Convolutional Neural Network (U-Net)
^j J48 Decision Tree (J48)
^k Maximum Entropy Model (Maxent)
^l Genetic Algorithm (GA).

Table 2. Performance of best prediction models from the research studies

Authors	Model	Acy(%)	Sny(%)	Spy(%)	AUC(%)	MSE ^a	R ^{2b}	MAE ^c	RMSE ^d
[37]	SVR ^e					0.19			
[26]	TAN ^f				89.00				
[34]	RFC ^g	82.60	60.00	96.60					
[12]	LASSO ^h	78.30	81.10	57.10					
[28]	SVM ⁱ	86.55			85.48				
[29]	GWR ^j					0.050	0.85	0.012	
[32]	SVR						0.75	0.44	
[35]	FFN ^k								0.67

Blank cells in the table indicate metrics that were not reported or utilized in the original studies.

^aMean Squared Error (MSE)

R-squared (R²)

^cMean Absolute Error (MAE)

^dRoot Mean Squared Error (RMSE)

^eSupport Vector Regression (SVR)

^fTree Augmented Naïve network (TAN)

^gRandom Forest classifier (RFC)

^hLeast Absolute Shrinkage and Selection Operator Regression (LASSO)

ⁱSupport Vector Machine (SVM)

^jGeographically Weighted Regression (GWR)

^kFeed Forward Network (FFN)

Datasets and Data Sources

ML and DL studies for leptospirosis prediction and diagnosis rely on diverse datasets with distinct characteristics (Table 3). Public datasets, typically sourced from government agencies, provide large-scale environmental and epidemiological data ideal for transmission modeling. For instance, the Thai Surveillance System [37] offered monthly rainfall measurements (0-450 mm range), soil pH values (4.5-8.2), and 30-meter resolution elevation data across five Southeast Asian countries from 2003-2018, comprising over 15,000 data points.

These datasets enabled regional risk prediction but lacked individual patient details. Similarly, Malaysia Meteorological Department records [30] provided 364 weeks of hourly rainfall data (0-65 mm/hr) and relative humidity (45-100%) paired with PCR-confirmed cases across 12 districts, demonstrating how high-resolution temporal data improves ANN-based outbreak forecasting.

Private clinical datasets, while smaller in scale, delivered granular patient-level information crucial for diagnostic accuracy. The Napoleón Franco Pareja Children's Hospital dataset [21] included 136 pediatric cases with detailed clinical parameters: fever duration (1-21 days), liver enzyme levels (AST 15-980 U/L), and urban/rural residence markers. More extensive Brazilian hospital records [22] encompassed 4,675 cases with serial creatinine measurements (0.2-9.8 mg/dL) and household GPS coordinates, though missing 12% of lab results. These datasets typically included three key data categories: (1) clinical biomarkers (serum creatinine, MAT titers 1:50-1:6400), (2) demographic information (age, gender in 89% of studies), and (3) epidemiological risk factors (rodent exposure indices).

Advanced studies combined multiple data types to overcome individual limitations. Research in China [23] integrated 2,741 CDC case reports with satellite-derived nighttime light indices (0-63 DN values) and river density maps (0-5.7 km/km²), achieving exceptional predictive performance (AUC 0.95-0.96). Hydrological studies in New Caledonia [35] correlated 226 water samples (turbidity 0-1,000 NTU, qPCR 10-10⁶ copies/mL) with 5-minute rainfall events, demonstrating how microenvironmental data enhances transmission understanding. These multimodal approaches compensated for individual dataset constraints through: (1) temporal complementarity (monthly climate + daily case reports), (2) spatial layering (watershed hydrology + village coordinates), and (3) clinical-environmental linkages (serum markers + livestock density).

Table 3. Comprehensive Dataset Characteristics of Included Studies

Authors	Data Type/Source	Data Categories	Sample Size	Temporal Resolution	Spatial Resolution
[37]	Public/ Thai surveillance system	• Environmental: Daily rainfall (mm), soil pH • Climatic: Max/min temperatures (°C) • Topographic: Elevation (SRTM 30m)	NS	Monthly 2003-2018	Regional (5 countries)
[30]	Public/ Malaysia Meteorologic al Dept	• Meteorological: Hourly rainfall (mm), RH (%) • Clinical: PCR-confirmed cases	364 weeks	Weekly 2011-2017	District-level (n=12)
[21]	Private/ Hospital records	• Clinical: Fever days, liver enzymes (U/L) • Demographic: Age, gender, urban/rural	136 patients	Single admission	Hospital catchment
[22]	Private/ Hospital records	• Clinical: Serum creatinine (mg/dL) • Epidemiological: Rodent exposure index • Geospatial: Residence coordinates	4,675 cases	2009-2016	Household-level
[24]	Private/ Gonçalo	• Clinical: Patient records • Epidemiological: Risk	4675 suspected	Retrospective (hospital)	Hospital/communit

	Moniz Institute (IGM), Federal University of Bahia (UFBA), Yale School of Public Health	factors, daily activities	cases (2046 confirmed, 2629 unconfirmed)	and community cohort)	y level (Salvador, Brazil)
[31]	Private/Medical records	- Clinical: Jaundice severity scale (0-3) - Laboratory: ELISA optical densities - Comorbidities: Diabetes status	800 patients	Retrospective 5y	Single tertiary center
[27]	Private/Bacterial cultures	- Genomic: 16S rRNA sequences - Proteomic: MALDI-TOF peaks (2k-20k m/z)	116 isolates	2015-2018	Lab-level
[26]	Private/Serosurvey + GIS	- Serological: MAT titers (1:50-1:6400) - Environmental: Livestock density/km² - Village attributes: Sanitation index	2,152 people	Dry/wet season	GPS coordinates (82 villages)
[34]	Private/Hydrological stations	- Streamflow (m³/s) - Water level (m) - Case reports (weekly)	517 weeks	Daily weekly →	Watershed-level
[12]	Private/Hospital EMR	- Vital signs: MAP (mmHg) - Labs: Creatinine (μmol/L) - Outcomes: Mortality	295 patients	2009-2022	3 hospitals
[28]	Mixed/MODIS + CDC	- Satellite: NDVI, LST (°C) - Case reports: District-level - Topography: Slope (%)	1,863 cases	Monthly 2009-2014	District-level
[33]	Public/Microscopy images	- Pixels: 256×256 RGB - Annotations: Spirochete masks	366 images	N/A	Pixel-level
[25]	Public/SINAN database	- Case reports: ICD-10 coded - Symptoms: 23-item checklist	890 cases	2007-2016	State-level

[29]	Public/ National Ministry of Health and Treatment of Iran, National Centre of Statistics of Iran, Meteorology Agency of Iran	• Disease: Positive ELISA test results	1186 positive cases (2009– 2011)	Longitudinal (2009–2011)	District- level (Gilan Province, Iran)
[32]	Public/ Island surveillance	• Climate: CHIRPS rainfall (mm) • Case counts: Ministry reports	Monthly NS	2010-2022	Island- level
[35]	Private/ Water monitoring	• Hydrological: Turbidity (NTU) • Microbiological: qPCR (copies/mL) • Weather: 5-min rainfall	226 samples	Event-based	3 km ² watershed
[23]	Public/ China CDC	• Environmental: River density (km/km ²) • Socioeconomic: Nightlight index • Cases: Lab-confirmed	2,741 cases	Annual 2004-2014	County- level

Not Specified (NS)

Centers for Disease Control and Prevention (CDC)

Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS)

Enzyme-Linked Immunosorbent Assay (ELISA)

Electronic Medical Records (EMR)

Geographic Information System (GIS)

Global Positioning System (GPS)

International Classification of Diseases 10th Revision (ICD-10)

Matrix-Assisted Laser Desorption/Ionization Time-of-Flight (MALDI-TOF)

Microscopic Agglutination Test (MAT)

Moderate Resolution Imaging Spectroradiometer (MODIS)

Normalized Difference Vegetation Index (NDVI)

Not Specified (NS)

Nephelometric Turbidity Units (NTU)

Polymerase Chain Reaction (PCR)

Quantitative Polymerase Chain Reaction (qPCR)

Relative Humidity (RH)

Red-Green-Blue color model (RGB)

Ribosomal Ribonucleic Acid (rRNA)

Sistema de Informação de Agravos de Notificação (SINAN)

Shuttle Radar Topography Mission (SRTM).

Standard units of measurement (mm, °C, m, etc.)

Discussion

This systematic review of 17 studies (2012–2024) shows that while ML and DL techniques achieve promising accuracy (80%–98%) in leptospirosis prediction and diagnosis, three critical limitations hinder clinical translation: (1) reliance on small, private clinical datasets limiting generalizability; (2) inconsistent validation methods, with only 11.8% employing temporal validation despite seasonality; and (3) underutilization of advanced techniques like transfer learning (0% adoption) and data augmentation (5.9% adoption).

The most effective algorithms varied by task—SVR and Random Forest for prediction, versus U-Net CNNs for microscopy-based diagnosis—but all models faced challenges related to data quality, sample size, and geographic bias.

Table 4. Summary of Strengths and Limitations Identified in ML and DL Studies for Leptospirosis Prediction and Diagnosis.

Strengths	Limitations
High predictive performance (80%–98% accuracy).	Small, private datasets limit generalizability.
Variety of ML/DL algorithms applied (e.g., RF, SVM, CNN).	Lack of external validation across datasets.
Integration of clinical and environmental data in some studies.	Underutilization of transfer learning and ensemble methods.
Growing research interest and recent publications.	Inconsistent evaluation metrics (accuracy, AUC, sensitivity).

While these methods show strong performance (80% - 98% accuracy in some cases), their real-world applicability remains limited due to dataset constraints, validation inconsistencies, and underutilization of advanced techniques such as transfer learning and ensemble learning. Addressing these gaps is essential to improve the robustness and clinical adoption of AI-driven leptospirosis diagnostics.

The most frequently used ML techniques for prediction tasks were SVR and Random Forest, while ANNs and CNNs, particularly U-Net, were commonly applied for diagnosis. This aligns with the increasing popularity of supervised ML methods in disease prediction.

Performance was primarily assessed using metrics such as accuracy, sensitivity, specificity, precision, and F1-Score, although AUC offers a more comprehensive measure of model performance, especially in binary classification tasks. Hybrid methods often produced better outcomes, with Random Forest and U-Net demonstrating strong accuracy and sensitivity in leptospirosis classification tasks. For instance, [31] achieved 87% accuracy and 91% sensitivity using Random Forest for diagnosis, suggesting that ensemble techniques could further enhance model performance.

Despite these advancements, there are still limitations in reaching clinician-level accuracy, especially when dealing with smaller datasets or limited training data. Acceptable performance thresholds varied, with some studies using an AUC score of 0.96 or higher as a benchmark, but this was not universally applied. Additionally, most studies relied on private, hospital-based datasets, limiting generalizability. The lack of external validation raises concerns about potential bias, emphasizing the need for public datasets and standardized validation protocols to improve cross-study comparability.

Cross-validation was the predominant method for evaluating the models, with k-fold or leave-one-out cross-validation being most common. However, due to the relatively small sample sizes in many studies, the conclusions drawn may not be as robust. Split validation (e.g., 80:20 or 70:30 splits) was also used, but inconsistencies across studies hindered direct model comparisons. Future work should prioritize establishing standardized protocols to enhance consistency and reliability in ML and DL evaluations.

One significant finding of this review was the absence of pre-trained models. Most studies developed models from scratch, limiting the generalizability and scalability of these models. Transfer learning involves using a pre-trained model as a starting point and fine-tuning it for a specific task. It has proven effective in various fields, particularly in image analysis and natural language processing, by significantly improving performance on tasks with limited data.

The lack of transfer learning in these studies suggests a potential area for future research, as it could enhance the performance of DL models in leptospirosis prediction and diagnosis, especially in cases where training data is limited. Similarly, data augmentation, which helps expand training datasets through transformations (e.g., rotations, translations, or noise), was only used in one study, highlighting a missed opportunity to improve model robustness.

Challenges and Research Gaps in ML and DL Applications for Leptospirosis Prediction and Diagnosis

While ML and DL have significant potential in leptospirosis research, this review also identified key findings that limit their broad application in clinical and public health settings.

Challenge 1: Limited Data Availability and Quality

One major challenge faced by the studies was the limited availability and quality of data. Many datasets, such as the one used in [22], had missing or incomplete data, which reduced the accuracy and generalizability of the models. Small sample sizes were a frequent issue as well, as observed in [21], where a dataset of only 136 patients resulted in high variability in model performance, especially for underrepresented cases like leptospirosis. Retrospective data collection posed further challenges, leading to biases in model training and evaluation, as noted by [31].

Challenge 2: Generalizability and Regional Bias

Several studies, such as [30,32], identified biases in data collection, such as under- and over-reporting, regional disparities, and sampling biases, which hindered the broader applicability of the models. Additionally, studies like [23] noted issues with the spatial resolution of environmental data, affecting the precision of fine-scale risk mapping. Moreover, the lack of external validation across different regions and datasets, as seen in studies like [12], increased the risk of overfitting and limited the broader applicability of findings.

The review showed inconsistent performance across different ML and DL algorithms. For example, models like SVM, ANN, and CNNs performed well, with accuracy reaching up to 98% [33]. However, advanced DL architectures like ResNet, Inception, and VGG were rarely utilized [27,32]. Similarly, none of the studies applied transfer learning, a technique that could enhance performance, particularly when data is scarce [12,30, 38]. Beyond data related challenges, there are also limitations in the ML/DL techniques currently applied to leptospirosis diagnosis, as discussed next.

Challenge 3: Underutilization of Advanced Techniques

Advanced ensemble techniques, such as XGBoost and Adaboost, were notably underutilized [28].

While Random Forest models and U-Net architectures performed well in specific tasks [31,33], ensemble methods could provide better predictive power when combining ML and DL approaches. The review also highlighted the minimal use of data augmentation techniques, which could help address the small sample size issues observed in many studies [33].

Limitations of this Systematic Review

While this systematic review provides valuable insights, it has certain limitations. The heterogeneity of study designs, dataset sizes, and performance metrics precluded a meta-analysis, limiting our ability to provide a standardized comparison of model performances. Additionally, the reliance on published studies may have introduced publication bias, as studies with less favorable results may have remained unpublished. Future systematic reviews should aim to standardize reporting metrics and ensure broader dataset accessibility to improve comparability across studies.

Recommendations for Future Research

Based on these findings, future research in ML and DL applications for leptospirosis should focus on the following areas:

- **Integration of Advanced DL Techniques:** Future studies should explore the potential of advanced DL architectures, such as ResNet and Inception, which are known to improve predictive performance, especially in image-based analysis.
- **Leveraging Pre-trained Models and Transfer Learning:** Research should investigate how pre-trained models can be fine-tuned for leptospirosis applications, particularly in data-limited scenarios.
- **Utilization of Ensemble and Hybrid Approaches:** Advanced ensemble techniques like XGBoost and hybrid ML-DL approaches should be explored to improve model accuracy and robustness.
- **Broader Geographic Representation:** Most studies focused on regions like Southeast Asia and Brazil, with limited research in other high-risk areas like Africa and Central America. Expanding research to these regions will improve model generalizability.
- **Addressing Socioeconomic and Environmental Factors:** Future models should incorporate important variables such as sanitation and income levels to better capture the complexities of leptospirosis transmission in real-world settings.

Addressing these areas will be critical for advancing robust, generalizable, and clinically impactful AI-based leptospirosis prediction models.

The primary goal moving forward is to aggregate a comprehensive dataset from diverse sources and develop a robust data library to enhance the accuracy and reliability of leptospirosis prediction models. Given the heterogeneity of data features across different studies—ranging from clinical records to environmental data—the focus will be on standardizing and harmonizing these features for better model integration. By consolidating larger and more varied datasets, we aim to improve model generalization and tackle current challenges related to small sample sizes and overfitting. This unified dataset will serve as a foundation for applying advanced techniques, such as transfer learning and ensemble methods, to further enhance the predictive power of ML and DL models in leptospirosis detection.

Conclusion

This systematic review examined ML and DL techniques for leptospirosis prediction and diagnosis by analyzing algorithm performance, evaluation methods, and challenges. While models such as SVM, ANN, Decision Trees, and CNNs have shown strong predictive power, most studies have

relied on private hospital-based datasets, limiting generalizability.

One key reason for the predominance of private datasets is that they often include detailed patient-level clinical information (e.g., laboratory values, comorbidities, and symptoms) essential for developing diagnostic models. In contrast, available public datasets mainly provide aggregated epidemiological or environmental data, which, while valuable for outbreak prediction, lack the granular patient-specific features necessary for individual diagnosis. As a result, limited use of public datasets reflects the inherent constraints in the nature and detail of publicly available data, rather than a preference by researchers.

Furthermore, the lack of advanced techniques like transfer learning and ensemble methods remains a concern, along with small sample sizes and inconsistent validation protocols. Overall, while significant progress has been made, there is considerable potential to improve the accuracy and generalizability of leptospirosis prediction models by integrating more comprehensive datasets and adopting advanced AI methodologies in future research.

Acknowledgment

This study was conducted under the leadership of Suhila Sawesi, who was responsible for the conceptualization, data extraction, analysis, and manuscript preparation and submission. Arya Jadhav assisted with all steps, including data extraction, analysis, and the manuscript preparation. Bushra Rashrash participated by assisting with data extraction, analysis, and providing critical input throughout the study. The authors declare no financial conflicts of interest related to this research or publication. No external sponsors were involved in the review, approval, or funding of this manuscript.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the research, authorship, and/or publication of this article.

Abbreviation List

ANN – artificial neural networks

AUC – area under the curve

CNN – convolutional neural networks

DL – deep learning

DT – decision trees

GA – genetic algorithm

GLM – generalized linear models

GWR – geographically weighted regression

JRIP – repeated incremental pruning to produce error reduction

KNN – k-nearest neighbors

MAE – mean absolute error

MAT – microscopic agglutination test

ML – machine learning

MLP – multilayer perceptron

MSE – mean squared error

NB – naive bayes

PCR – polymerase chain reaction

PRISMA – Preferred Reporting Items for Systematic Reviews and Meta-Analyses

RDT – rapid diagnostic tests

RMSE – root mean squared error

SVM – support vector machines

Legends for Figures

Figure 1. PRISMA flow diagram of the search and selection process for identifying relevant studies. *This figure outlines the flow of information through the different phases of the systematic review.*

Figure 2. Distribution of studies by publication year and prediction/diagnosis tasks. *This figure highlights the increasing trend in the application of ML/DL to leptospirosis over time.*

Figure 3. Datasets classified by study type (prediction or diagnosis).

This figure shows how datasets are distributed based on their use in prediction or diagnosis tasks.

Figure 4. Heatmap showing the distribution of machine learning classifiers across different studies. *This figure demonstrates which machine learning classifiers are most commonly used for leptospirosis tasks.*

References

1. Leptospirosis: Fact sheet. World Health Organization. January 1, 2009. Accessed August 9, 2024. <https://www.who.int/publications/i/item/B4221>.
2. About leptospirosis. Centers for Disease Control and Prevention. June 24, 2024. Accessed August 9, 2024. <https://www.cdc.gov/leptospirosis/about/index.html>.
3. 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*. 2023;2:100045. doi:10.1016/j.soh.2023.100045
4. Ching T, Himmelstein DS, Beaulieu-Jones BK, et al. Opportunities and obstacles for deep learning in biology and medicine. *Journal of The Royal Society Interface*. 2018;15(141):20170387. doi:10.1098/rsif.2017.0387
5. Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in Healthcare. *Nature Medicine*. 2019;25(1):24-29. doi:10.1038/s41591-018-0316-z
6. Topol EJ. High-performance medicine: The convergence of human and Artificial Intelligence. *Nature Medicine*. 2019;25(1):44-56. doi:10.1038/s41591-018-0300-7
7. Angermueller C, Pärnamaa T, Parts L, Stegle O. Deep Learning for Computational Biology. *Molecular Systems Biology*. 2016;12(7). doi:10.15252/msb.20156651
8. Valente M, Bramugy J, Keddie SH, et al. Diagnosis of human leptospirosis: Systematic review and meta-analysis of the diagnostic accuracy of the *Leptospira* microscopic agglutination test, PCR targeting LFB1, and IGM Elisa to *Leptospira* fainei serovar Hurstbridge. *BMC Infectious Diseases*. 2024;24(1). doi:10.1186/s12879-023-08935-0
9. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436-444. doi:10.1038/nature14539
10. Litjens G, Kooi T, Bejnordi BE, et al. A survey on Deep Learning in medical image analysis. *Medical Image Analysis*. 2017;42:60-88. doi:10.1016/j.media.2017.07.005

11. Sarvamangala DR, Kulkarni RV. Convolutional neural networks in medical image understanding: A survey. *Evolutionary Intelligence*. 2021;15(1):1-22. doi:10.1007/s12065-020-00540-3
12. Galdino GS, de Sandes-Freitas TV, de Andrade LG, et al. Development and validation of a simple machine learning tool to predict mortality in leptospirosis. *Scientific Reports*. 2023;13(1). doi:10.1038/s41598-023-31707-4
13. Moher D, Liberati A, Tetzlaff J, Altman DG, Prisma Group T. Preferred reporting items for systematic reviews and meta-analyses: The Prisma statement. *PLoS Medicine*. 2009;6(7). doi:10.1371/journal.pmed.1000097
14. Moons KG, de Groot JA, Bouwmeester W, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: The charms checklist. *PLoS Medicine*. 2014;11(10). doi:10.1371/journal.pmed.1001744
15. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *Journal of Clinical Epidemiology*. 2019;110:12-22. doi:10.1016/j.jclinepi.2019.02.004
16. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159. doi:10.2307/2529310
17. McHugh ML. Interrater Reliability: The kappa statistic. *Biochemia Medica*. Published online October 15, 2012:276-282. doi:10.11613/bm.2012.031
18. Mu Y, He D. The potential applications and challenges of chatgpt in the medical field. *International Journal of General Medicine*. 2024;Volume 17:817-826. doi:10.2147/ijgm.s456659
19. Miao J, Thongprayoon C, Fülöp T, Cheungpasitporn W. Enhancing clinical decision-making: Optimizing Chatgpt's performance in hypertension care. *The Journal of Clinical Hypertension*. 2024;26(5):588-593. doi:10.1111/jch.14822
20. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (tripod): The tripod statement. *Annals of Internal Medicine*. 2015;162(1):55-63. doi:10.7326/m14-0697
21. Caicedo Torres W, Quintana M, Pinzón H. Differential diagnosis of hemorrhagic fevers using ARTMAP and an Artificial Immune System. *International Journal of Artificial Intelligence*. 2013;11(13 A):150-169. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84879759875&partnerID=40&md5=e126f6afb294df43dee6cf023435ca5a>
22. Nery NR, Claro DB, Lindow JC. Classification model analysis for the prediction of leptospirosis cases. *2016 11th Iberian Conference on Information Systems and Technologies (CISTI)*. Published online June 2016:1-6. doi:10.1109/cisti.2016.7521494
23. Zhao J, Liao J, Huang X, et al. Mapping risk of leptospirosis in China using environmental and Socioeconomic Data. *BMC Infectious Diseases*. 2016;16(1). doi:10.1186/s12879-016-

1653-5

24. Nery NR, Claro DB, Lindow JC. Prediction of leptospirosis cases using classification algorithms. *IET Software*. 2017;11(3):93-99. doi:10.1049/iet-sen.2016.0193
25. Lopez DM, de Mello FL, Giordano Dias CM, et al. Evaluating the surveillance system for spotted fever in Brazil using machine-learning techniques. *Frontiers in Public Health*. 2017;5. doi:10.3389/fpubh.2017.00323
26. Mayfield HJ, Smith CS, Lowry JH, et al. Predictive risk mapping of an environmentally-driven infectious disease using spatial Bayesian Networks: A case study of leptospirosis in Fiji. *PLOS Neglected Tropical Diseases*. 2018;12(10). doi:10.1371/journal.pntd.0006857
27. Sonthayanon P, Jaresitthikunchai J, Mangmee S, et al. Whole cell matrix assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) for identification of *Leptospira* spp. in Thailand and Lao PDR. *PLOS Neglected Tropical Diseases*. 2019;13(4). doi:10.1371/journal.pntd.0007232
28. Ahangarcani M, Farnaghi M, Shirzadi MR, Pilesjö P, Mansourian A. Predictive risk mapping of human leptospirosis using support vector machine classification and Multilayer Perceptron Neural Network. *Geospatial Health*. 2019;14(1). doi:10.4081/gh.2019.711
29. Mohammadinia A, Saeidian B, Pradhan B, Ghaemi Z. Prediction mapping of human leptospirosis using ann, GWR, SVM and GLM approaches. *BMC Infectious Diseases*. 2019;19(1). doi:10.1186/s12879-019-4580-4
30. Rahmat F, Zulkafli Z, Juraiza Ishak A, Mohd Noor SB, Yahaya H, Masrani A. Exploratory Data Analysis and artificial neural network for prediction of leptospirosis occurrence in Seremban, Malaysia based on Meteorological Data. *Frontiers in Earth Science*. 2020;8. doi:10.3389/feart.2020.00377
31. Shenoy S, Rajan AK, Rashid M, et al. Artificial Intelligence in differentiating tropical infections: A step ahead. *PLOS Neglected Tropical Diseases*. 2022;16(6). doi:10.1371/journal.pntd.0010455
32. Douchet L, Goarant C, Mangeas M, Menkes C, Hinjoy S, Herbreteau V. Unraveling the invisible leptospirosis in Mainland Southeast Asia and its fate under climate change. *Science of The Total Environment*. 2022;832:155018. doi:10.1016/j.scitotenv.2022.155018
33. Kulkarni A, Sai Dinesh Reddy P, Bassi R, Kashyap SK, Vijayalakshmi M. Spirochaeta bacteria detection using an effective semantic segmentation technique. *Advances in Intelligent Systems and Computing*. Published online 2022:355-365. doi:10.1007/978-981-16-9573-5_26
34. Jayaramu V, Zulkafli Z, De Stercke S, et al. Leptospirosis modelling using hydrometeorological indices and Random Forest Machine Learning. *International Journal of Biometeorology*. Published online January 31, 2023. doi:10.1007/s00484-022-02422-y
35. Thibeaux R, Genthon P, Govan R, et al. Rainfall-driven resuspension of pathogenic *Leptospira* in a leptospirosis hotspot. *Science of The Total Environment*. 2024;911:168700.

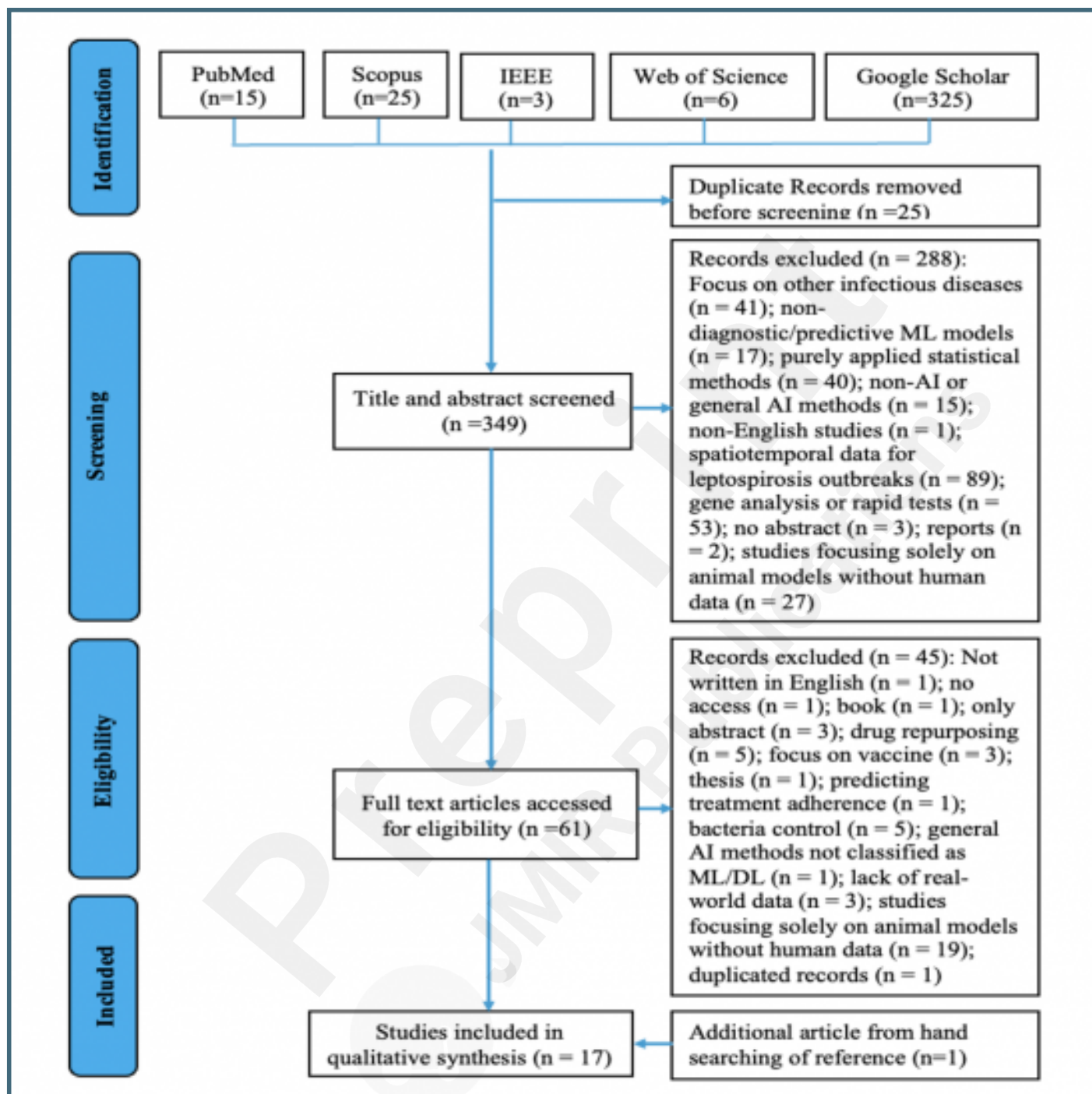
doi:10.1016/j.scitotenv.2023.168700

36. Nakano R, Oyamada Y, Ozuru R, Miyahara S, Yoshimura M, Hiromatsu K. A Deep Learning-Based Evaluation for Microscopic Agglutination Test. *bioRxiv*. Published online 2024:2024-04
37. Douchet L, Menkes C, Herbreteau V, et al. Climate-driven models of leptospirosis dynamics in Tropical Islands from three oceanic basins. *PLOS Neglected Tropical Diseases*. 2024;18(4). doi:10.1371/journal.pntd.0011717
38. Jadhav A, Sawesi S, Rashrash B. Bias and Generalizability Challenges in Machine Learning Models for Leptospirosis. In 2024 IEEE International Conference on Big Data (BigData) 2024 Dec 15 (pp. 4989-4995). IEEE. 10.1109/BigData62323.2024.10825588

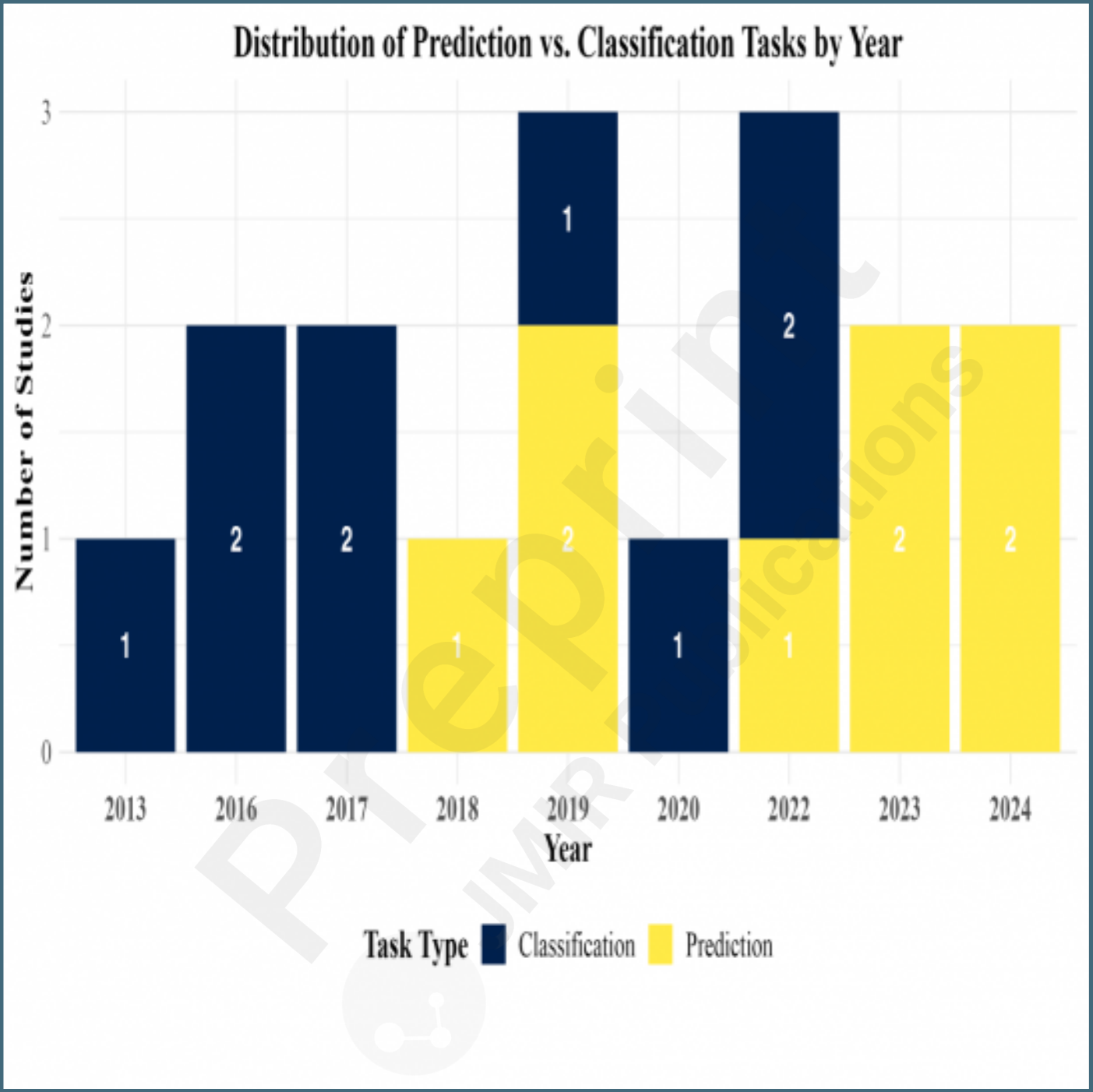
Supplementary Files

Figures

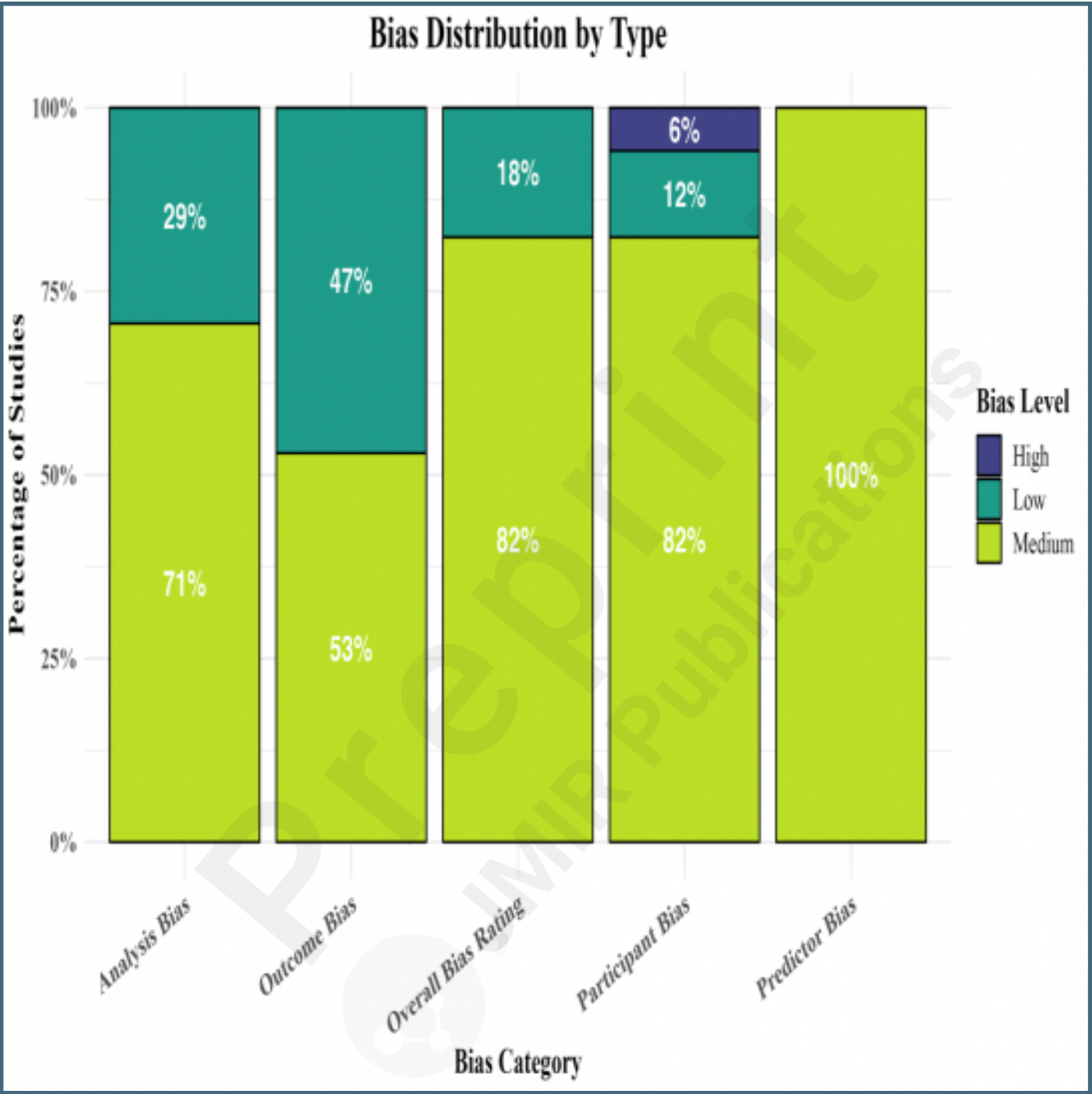
PRISMA flow diagram of the search and selection process for identifying relevant studies.



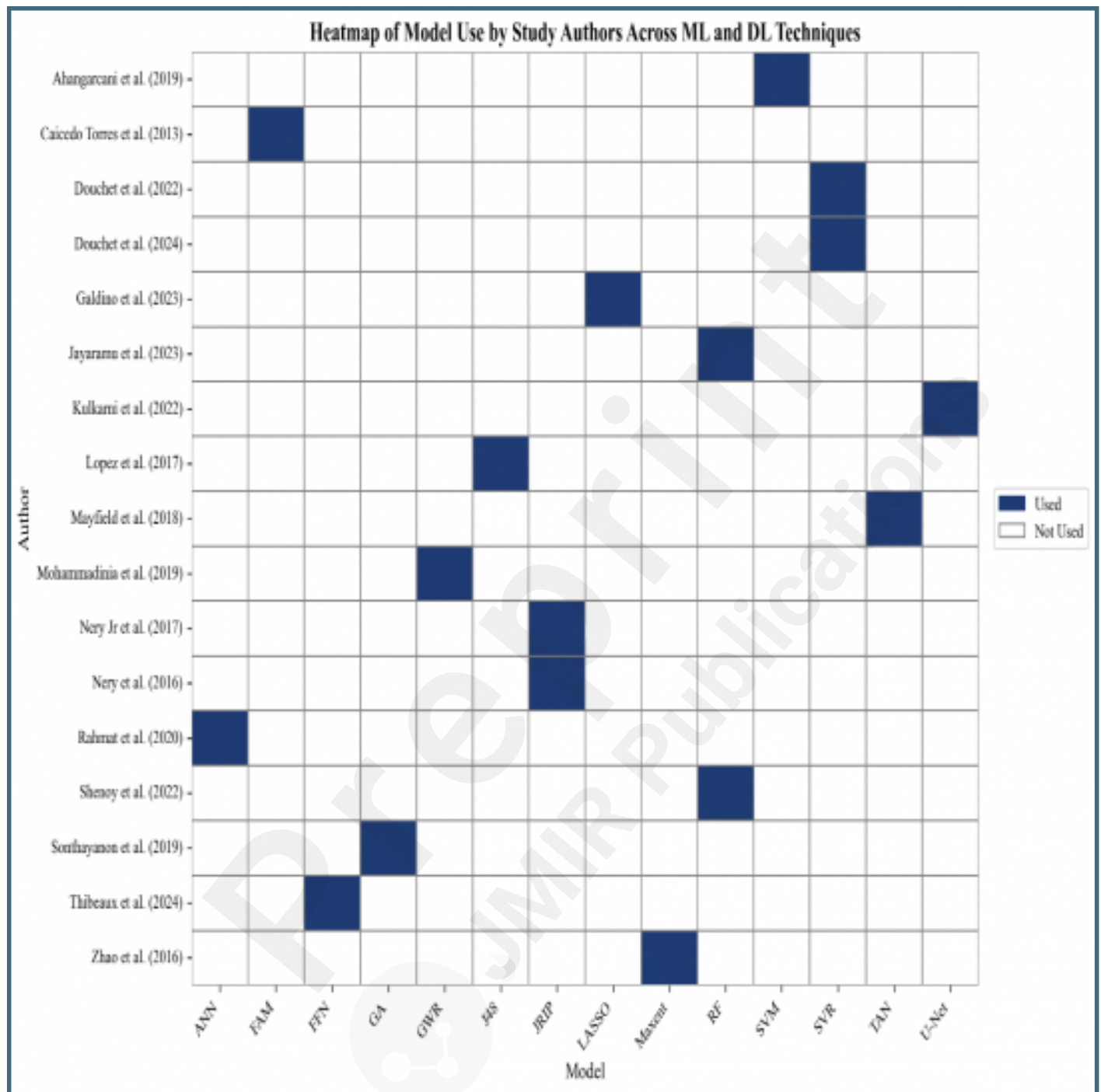
Distribution of Studies on Machine Learning and Deep Learning Applications for Leptospirosis Diagnosis and Prediction by Year and Task Type.



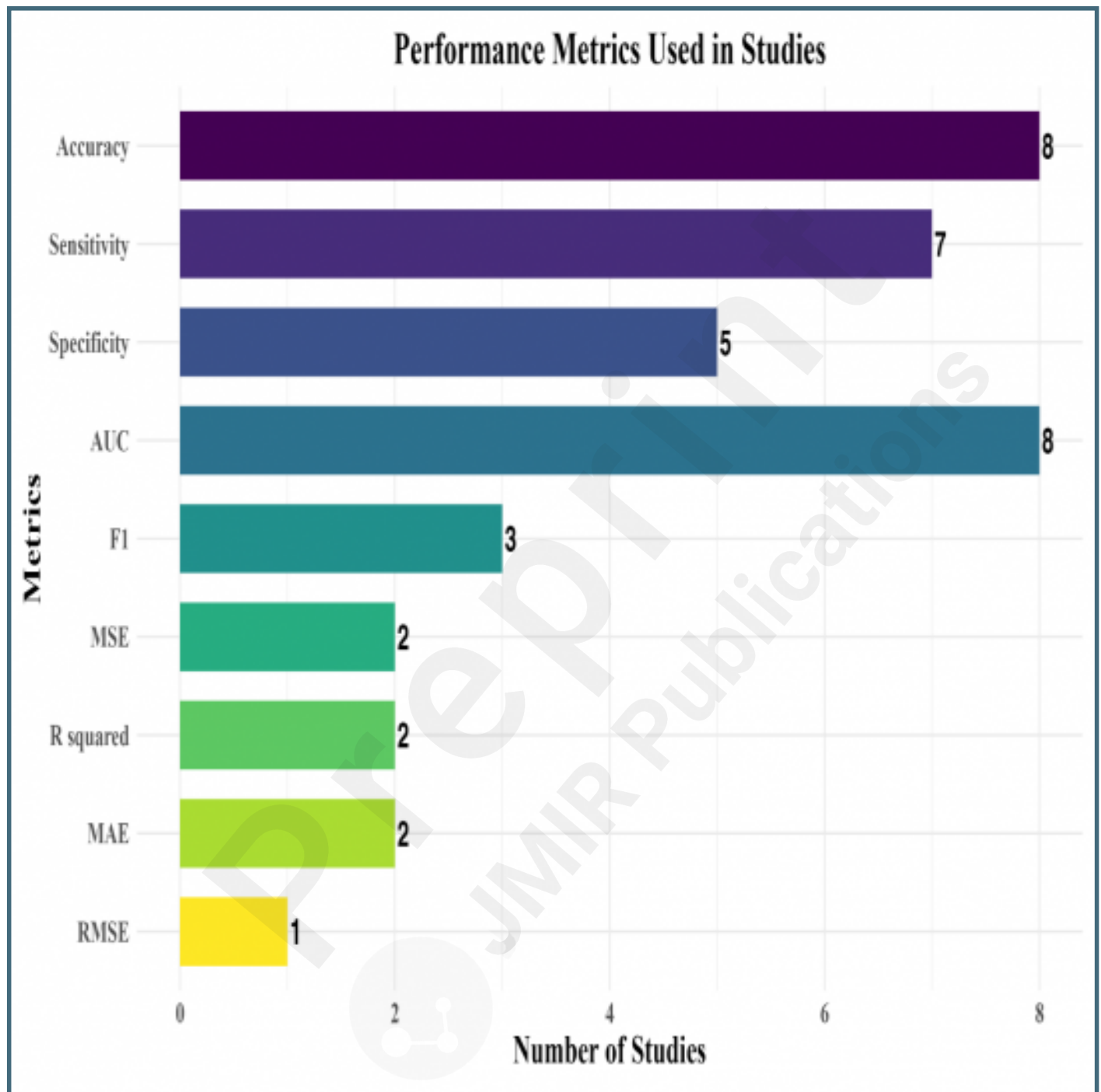
Distribution of Risk of Bias Across Domains in Machine Learning and Deep Learning Studies for Leptospirosis.



Heatmap of Classifier Usage Across Included Studies.



Number of Algorithm Performance Metrics Used in Reviewed Articles of Dataset Types Used (Public/Private).



Multimedia Appendixes

Search Terms.

URL: <http://asset.jmir.pub/assets/3f15947cf81f30e08915d9bc6312eaf7.docx>

Agreement Between Human Reviewers and ChatGPT-4o in Study Screening.

URL: <http://asset.jmir.pub/assets/1ada47b1613ae7a7425b329ed6acd700.docx>

Quality Assessment/Risk of Bias.

URL: <http://asset.jmir.pub/assets/4cc63c55460065108b845a96ccf33eb5.docx>

Characteristics of Included Studies.

URL: <http://asset.jmir.pub/assets/3c1f974be0d4d865fee68381ffbd4932.docx>

PRISMA 2020 Checklist for Systematic Reviews, detailing the sections, topics, checklist items, and their corresponding locations within the review.

URL: <http://asset.jmir.pub/assets/08fcbae4fc0275b1ca422c59b219cd9d.docx>