Published by

British Journal of Contemporary Research

BEXFORD PUBLISHING

Machine Learning for Pre-Test Malaria Diagnosis from Symptoms, Clinical Text and Speech: A PRISMA Systematic Review

Review Article | Volume 1 | Issue 6 | 2026 | Article Number: 070

Accepted: 03 November 2026 | Published: 10 November 2026 | ISSN: 2979-8582 (Online)

Crossref : **https://doi.org/10.67693/BJCR-6RSBUMFE**

Olalekan Akinbosoye Okewale[1], Olufemi S. Ojo[1], F. W. Ipeayeda[1], Segun Kazeem Fatoki[1], Mayowa Oyedepo Oyediran[1]

[1]Department of Computer Science, Ajayi Crowther University, Oyo, Nigeria

OAO 0000-0003-2272-8660, OSO 0000-0001-7505-7990, FWI 0009-0009-6055-9651, SKF 0009-0007-1921-1798, MOO 0000-0001-6756-1476

Correspondence: Olalekan Akinbosoye Okewale, okewale.akinbosoye@gmail.com

Abstract

In the absence of microscopy and rapid tests, malaria diagnosis is presumptive in many clinics in Africa, leading to overuse of drugs and masking other causes of fever. It has been proposed, based on pre-test data (e.g., symptoms, clinical notes, patient voice), to use standard ML models as triage tools. We searched PubMed and Scopus for studies from 2010 to 2026 that trained and evaluated standard ML models on pre-test clinical data to paint a high RQ1 for malaria diagnosis or on voice or clinical text for other infectious diseases (RQ1) and followed PRISMA 2020. We included 146 studies (RQ1=43; RQ2=103) out of 2598 studies. A modified PROBAST assessed the risk of bias. No malaria studies used speech, and only four used free text or language models. Excluding laboratory reference standard, 27 (63%) out of 43 studies for malaria employed 95% or greater accuracy, found in 14 out of 17 studies. Those with laboratory labelling (and independent adjudications) found discrimination modest, with AUCs between 0.60–0.95. However, chance-level predictions were observed when new speakers, recording devices, and entirely new datasets were used; only 14 of 103 studies mention independent data. These data justify pre-test malaria triage only if subsequent models include parasitological classification, patient-level speaker-independent splits, symptom-matched controls, external validation, and combined and separate reporting from each input.

Keywords: Clinical Decision Support, Malaria, Machine Learning, Natural Language Processing, Speech Analysis, Systematic Review

© 2026 The Authors. Published by Bexford Publishing Ltd

I. INTRODUCTION

Malaria is one of the leading infectious diseases contributing to morbidity and mortality in sub-Saharan Africa. In 2024, there were an estimated 282 million malaria cases and 610,000 deaths, with 95% of these deaths occurring in the WHO African Region (World Health Organisation, 2025b). The burden of the disease is only in one country, Nigeria (World Health Organisation, 2025b). The World Health Organisation recommends that all suspected cases be confirmed by microscopy or rapid diagnostic tests (RDTs) before treatment (World Health Organisation, 2025a). However, in practice, such testing is often limited (due to lack or shortage of the tests themselves) and many health workers just manage febrile patients based on clinical symptoms. Such presumptive treatment is a common practice in Nigerian health facilities and leads to over-reliance on artemisinin-based treatments and ignores many other causes of fever (Anjorin et al., 2023).

Machine learning presents an opportunity of assisting care providers in determining the actual cause of malaria when they attend to their patients before undergoing diagnosis. The model described may utilise the symptoms patient report, as well as notes on what clinical staff observe most often, at first patient contact, to score patients for the likelihood of malaria. Moreover, the approach may be useful in low-resource settings where laboratory tools are not always available, especially if features are extracted from speech recordings and unstructured clinical notes. Previous reviews have focused on either blood smear images or on NLP applications to public health in Africa; none have critically examined models based on clinical information prior to testing or addressed whether there is experience with speech recognition models for other infectious diseases.

The aims of this review are: RQ1: Identify and evaluate ML models developed and tested for malaria diagnosis using pre-parasitological clinical information (structured and free text) and assess how much they have been validated; RQ2: Investigate what can be learned from ML models using speech or text analysis for other infections (with emphasis on validation, as no speech model for malaria was expected). The rest of the paper reviews related work, describes the research methodology used in this work, provides the results and discussion, and finally concludes with the achievements of this work.

The contributions of this work are:

i. a PRISMA-implemented systematic review of 146 studies and an adapted PROBAST risk-of-bias assessment;
ii. evidence that reported accuracy is label- and validation-dependent, not just model classification-dependent;
iii. guidance on core design principles and reporting standards for developing effective multimodal triage strategies for pre-testing prioritisation in malaria.

II. RELATED WORKS

Malaria Diagnosis and the Decision before Testing

Microscopy of stained blood films remains the reference method for malaria diagnosis, and RDTs that detect parasite antigens have widened access in primary care (World Health Organisation, 2025a). Both depend on staff, supplies and laboratory capacity that are often missing in rural African clinics, and in Nigeria many febrile patients still receive antimalarials without a test (Anjorin et al., 2023). Clinical algorithms such as the Integrated Management of Childhood Illness guide decisions when testing is limited, but symptoms such as fever, headache and chills are shared with typhoid, viral infections and other causes of fever, so symptom-based rules have limited specificity (McLaughlin et al., 2022; Taneja et al., 2021). This is the decision that pre-test prediction models aim to improve.

© 2026 The Authors. Published by Bexford Publishing Ltd

Machine Learning for Malaria

Most malaria ML studies have conducted research on images of blood smears. Convolutional networks now classify infected and uninfected cells with very high accuracy on public datasets (Ahamed et al., 2025; Boit & Patil, 2024; Hoyos & Hoyos, 2024), and reviews have described the move from these unimodal image models to multimodal approaches (Amrevuawho et al., 2025). But image models need a smear and a microscope so that they only support the test rather than the decision before it. A smaller body of work predicts malaria as a result of symptoms, demographic data or routine laboratory values, typically using logistic regression, tree ensembles or shallow neural networks (Balerdi-Sarasola et al., 2024; Nyengera et al., 2026; Ojurongbe et al., 2023). There has been discussion of opportunities and risks with AI in African malaria programmes, but there has been a lack of assessment of the validity of AI models (Obeagu & Ogenyi, 2025).

Speech as a Health Signal

Speech reflects the respiratory, laryngeal and articulatory systems, and acoustic features such as Mel-frequency cepstral coefficients have long been used for pathological voice analysis. During the COVID-19 pandemic, crowdsourced databases of voice, cough and breath recordings were released (Sharma et al., 2020) and international challenges benchmarked detection systems (Coppock et al., 2023; Muguli et al., 2021). These efforts produced the largest body of evidence on speech for infectious disease, together with clear warnings about confounding: a controlled analysis of a large PCR-linked dataset found that audio classifiers did not outperform a simple symptom checker (Coppock et al., 2024). No comparable work exists for malaria.

Clinical Text and Language Models

Clinical notes record the presenting complaint, history and examination in free text. Deep learning on electronic health records (Shickel et al., 2018) and NLP for clinical decision support (Eguia et al., 2024) have grown rapidly, and multimodal large language models are now being applied to clinical text (AlSaad et al., 2024). In Africa, NLP for health is limited by scarce digital records and few resources for local languages (Hu et al., 2025). A specific risk for diagnostic models is that notes written after a test may contain the result or the diagnosis, which turns prediction into information retrieval (Kapoor & Narayanan, 2023).

Multimodal Fusion, Sequence Models and Optimisation

Multimodal machine learning combines heterogeneous inputs by early fusion of raw features, intermediate fusion of learned representations or late fusion of predictions (Baltrusaitis et al., 2019). Reviews in health care report that intermediate fusion is the most common choice and that gains over the best single modality are frequent but not guaranteed (Acosta et al., 2022; Kline et al., 2022; Stahlschmidt et al., 2022). Recurrent networks such as the long short-term memory (Hochreiter & Schmidhuber, 1997) and the gated recurrent unit (GRU) (Cho et al., 2014), often with attention (Bahdanau et al., 2015), are used for sequential speech and text. Metaheuristic algorithms are widely used for feature selection and hyperparameter tuning in disease diagnosis (Kaur, Kumar, et al., 2023), but their advantage over random search at an equal computational budget is rarely tested (Bergstra & Bengio, 2012; Bischl et al., 2023).

Methodological Standards for Prediction Models

PROBAST assesses the bias risk on prediction models (Wolff et al., 2019) and TRIPOD+AI defines a standard of reporting for machine learning models (Collins et al., 2024). Both frameworks stress the need for a reliable reference standard, ensuring predictors are assessed without prior knowledge of the outcomes, an adequate sample size and data used independently. Commonly, one of the reasons for the

© 2026 The Authors. Published by Bexford Publishing Ltd

poor results in machine learning is the contamination between training data and test data (Kapoor & Narayanan, 2023). These standards are the basis for the analysis in this review.

Research Gap

Existing reviews fail to view pre-test malaria models as a single category, do not address the evaluation process of their labels and validation, and do not explore speech or clinical text in malaria. Since speech and text models for malaria are not likely to be common, it is necessary to draw on the evidence from other infections to assess possible challenges. This review aims to fill these gaps.

III. METHODOLOGY

Protocol and Reporting

The evaluation was performed and described following PRISMA 2020 guidelines (Page et al., 2021). A process was developed for research questions, eligibility criteria, search terms, and data extraction forms before screening.

Eligibility Criteria

Table 1 summarises the eligibility criteria. Any research design was considered if it involved training or evaluating a ML or language model using data from actual patients, accompanied by reported diagnostic performance metrics. For research question 1 (RQ1), we used malaria as the target condition and needed to have the inputs to perform a parasitological test before the diagnosis and included symptoms, signs, demographic information, exposure history and routine haematology results as well as free text or speech data. No models that are solely based on images of blood smears were included in that group. For research question 2 (RQ2), we used any infectious disease as the target condition and the input was either voice communication or clinical text from a patient. We excluded input limited exclusively to cough sounds, breath sounds or lung auscultation (not images) as they do not contain spoken or written language that can be used to build a malaria triage tool. We also excluded reviews and editorials; descriptions of datasets lacking associated models; retracted publications; synthetic patient scenarios; and research focused primarily on information extraction rather than accurate diagnoses.

Table 1: Eligibility Criteria

Criterion	RQ1 (malaria)	RQ2 (other infections)
Population	Patients assessed for malaria	Patients assessed for an infectious disease
Index model	ML or language model using pre-test information: symptoms, signs, demographics, exposure, routine haematology, text or speech	ML or language model using speech or voice, or clinical or patient text, alone or with clinical data
Reference	Any stated diagnosis (laboratory reference recorded for appraisal)	Any stated diagnosis
Outcome	Diagnostic performance	Diagnostic performance
Excluded	Smear images only; synthetic data; reviews; no model	Cough, breath or images only; extraction-only outcomes; synthetic data; reviews
Period	2010 to 2026	2010 to 2026

© 2026 The Authors. Published by Bexford Publishing Ltd

Information Sources and Search

PubMed was searched for RQ1, and Scopus for RQ1 and RQ2, for records published from 2010 to 2026. Each search combined four concept blocks: the disease (malaria or Plasmodium for RQ1; infectious, febrile, respiratory and named infections for RQ2), the method (machine learning, deep learning, neural networks, NLP), the task (diagnosis, classification, prediction, detection) and the input (clinical, symptoms, records, text, speech, voice, audio, multimodal). The full strings are given in the supplementary material. No language or document-type filters were applied at search time.

Selection Process

Delineation of entries was made by DOI first, then PubMed identifier and finally by standardised title. The proceedings front matter, reviews and editorials, notes, correspondence, errata, retracted items and other non-English records were checked by rule-based filters. Screening of titles and abstracts was done directly by the authors. The primary author and the secondary author evaluated the screening process and made the decisions. The primary author recorded different decisions in the screening workbook, and any discrepancies were discussed and recorded. The correlation between the primary author and the initial pass was 97.8% (Cohen's kappa 0.85 with unmarked rows given as agreement).

Data Items and Risk of Bias

For each study included in our analysis, we extracted the study setting and country, data sources, samples, inputs to be used for modelling, reference standards (labels), models to be selected, and feature selection and optimisation. We also discussed the validation design, the imbalanced class problem, and the reported best results for all the studies. Using PROBAST (Wolff et al., 2019), we rated biases in four domains: participants (e.g., crowdsourced vs. case-control/case-report samples); predictors and labels (e.g., labels based on symptoms vs. based on test results, labelled as predictive vs. not predictive); and analytical systems (e.g., random split vs. resampling before splits). Each aspect was graded as low, moderate, or high risk of bias.

Synthesis

The datasets and targets were too heterogeneous to be considered for meta-analysis. Results were collected in a narrative style and summarised in tables according to input type, reference standard and validation design. For RQ2, studies were also considered to have strong validation if they used participant-disjoint, speaker-independent, temporal, cross-dataset or external evaluation.

IV. RESULTS AND DISCUSSION

Study Selection

A total of 2,598 records were identified (see Fig. 1). After excluding 306 duplicates and 498 reports that were ineligible by type or language, the remaining 1,794 records had their titles and abstracts screened. Two hundred reports were sought, with twenty-two not found. The remaining 178 were reviewed in detail, of which 32 were excluded due to ineligible inputs (coughs, images, or EEG data alone; n = 9), use of synthetic data sets (n = 6), poor reporting (n = 6), non-diagnostic (n = 6), lack of original model (n = 4), or withdrawal from publication (n = 1). Thus, there are 146 studies that were considered in this review: 43 for RQ1 and 103 for RQ2. The complete extraction table is available in the supplementary materials.

© 2026 The Authors. Published by Bexford Publishing Ltd

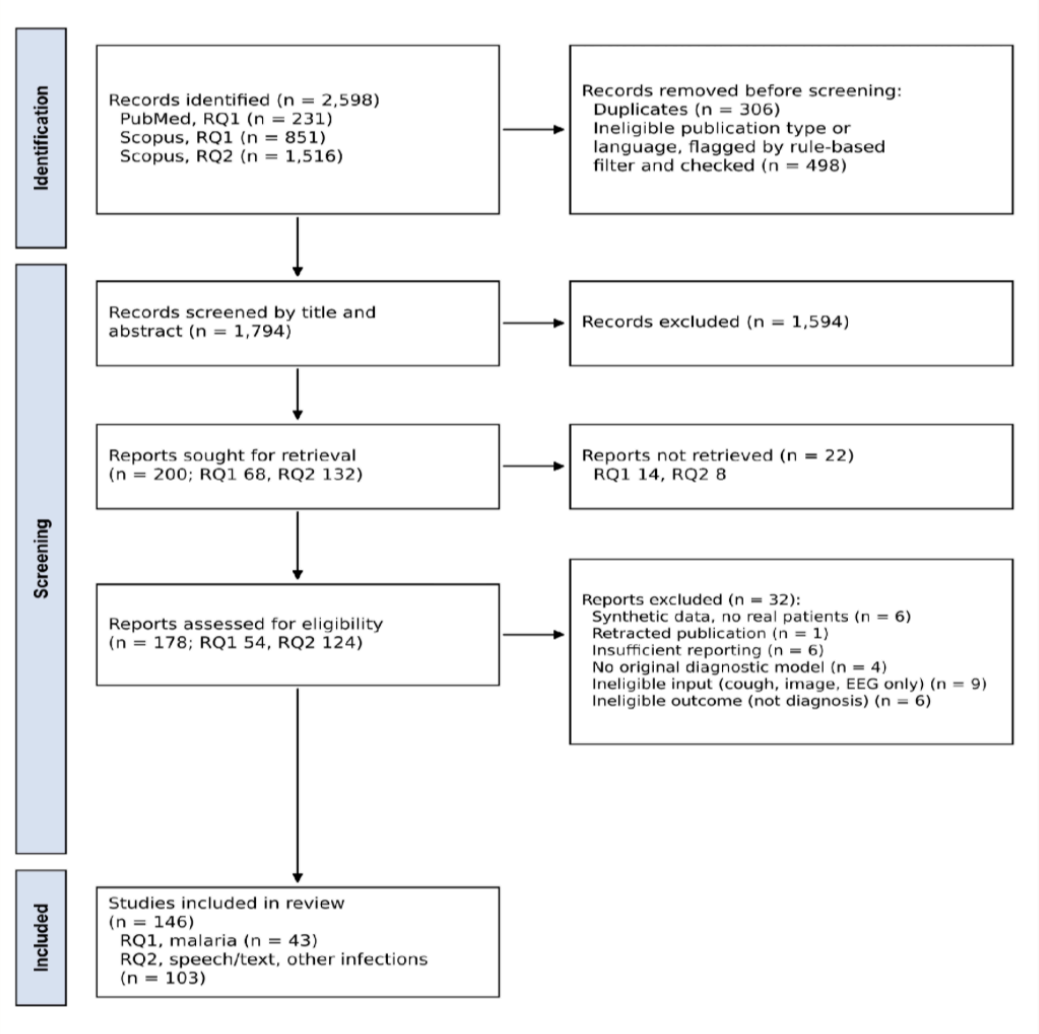


Figure 1: PRISMA 2020 flow diagram of study selection

RQ1: Characteristics of Malaria Studies

Table 2 summarises the 43 malaria studies. The field is recent and concentrated in a few countries: 42 studies were published from 2019 onwards and 23 from 2024, and 18 used Nigerian data. Thirty-three used structured symptoms and clinical signs, five used routine laboratory or haematology values, three used household or survey variables and four used free text or a language model. None used speech. One study described a fusion of voice, text, image and video features with recurrent networks but reported no data source, labels or absolute performance (Achary et al., 2025). Several datasets were reused across papers: a 337-patient Nigerian dataset appeared in four studies (Awe et al., 2025; Ayalew et al., 2024; Deepika & Pradeep Kumar, 2025; Okagbue et al., 2021) and a febrile-illness dataset from the Niger Delta in three (Amannah et al., 2025; Attai et al., 2024; Attai et al., 2025), so the 43 studies represent fewer independent samples than their number suggests.

Table 2: Characteristics of the 43 Malaria (RQ1) Studies

Characteristic	Studies
Published 2019 to 2023 / 2024 to 2026	42 / 23
Data from Nigeria / India / other Africa	18 / 10 / 8
Data from Asia / Latin America	4 / 3

© 2026 The Authors. Published by Bexford Publishing Ltd

Inputs: structured symptoms and signs	33
Inputs: routine laboratory or haematology	5
Inputs: household or survey variables	3
Inputs: free text or language model	4
Inputs: speech	0
Laboratory reference (microscopy, RDT, PCR, QBC)	16
Clinician, dataset or unstated label	27
Reported accuracy of 95% or more	17 (3 with laboratory reference)
Independent (external) evaluation	2
Overall risk of bias: low / moderate / high	1 / 17 / 25

RQ1: Reference Standards and Label Validity

Only 16 studies labelled malaria with microscopy, an RDT, PCR or a quantitative buffy coat test. The other 27 used a clinician's diagnosis, a label inherited from a public dataset, or did not state how malaria was confirmed. In some studies the label could be reconstructed from the inputs. Three studies from one Nigerian hospital trained models to predict likelihood classes that clinicians had assigned from the same symptoms, without laboratory confirmation, and reported detection rates of 97 to 98% (Oguntimilehin et al., 2019; Oguntimilehin, Babalola, et al., 2024; Oguntimilehin, Berida, et al., 2024). One study included the laboratory result among the predictors and reported 100% accuracy (Atoyebi et al., 2024), and another included diagnostic codes and a risk score and also reported 100% (Padmakala, 2026). Of the 17 studies that reported accuracy of 95% or higher, only three used a laboratory reference.

RQ1: Performance under Rigorous Evaluation

Studies with laboratory labels and careful validation told a different story. A Bayesian network built from 3,441 paediatric encounters in a national survey of Malawian facilities, with mRDT labels, reached an area under the curve (AUC) of 0.60 (Taneja et al., 2021). In Nigerian children, a tree ensemble improved on the Integrated Management of Childhood Illness algorithm but reached a sensitivity of 60% and a specificity of 79% against mRDT (McLaughlin et al., 2022). A penalised logistic regression from four Nigerian primary health centres, with nested cross-validation and feature selection inside the folds, reported a test AUC of 0.83 with a 95% confidence interval of 0.63 to 1.00, reflecting 35 positive cases (Ojurongbe et al., 2023). Higher discrimination was reported when routine laboratory values were added, for example an AUC of 0.92 to 0.95 in febrile travellers (Balerdi-Sarasola et al., 2024) and 0.866 for haematology parameters in Ghana (Morang'a et al., 2020), and when a haematology analyser's malaria flags were evaluated externally against blinded microscopy, with an AUC of 0.90 for vivax malaria (Dharap & Raimbault, 2020). Only two studies evaluated a model on data fully independent of its development (Dharap & Raimbault, 2020; Yelin et al., 2025). Table 3 lists the 16 studies with a laboratory reference standard.

Table 3: Malaria Studies with a Laboratory Reference Standard

Study	Setting, sample	Inputs	Best result
Taneja et al. (2021)	Malawi, 3,441	Symptoms, signs	AUC 0.60
McLaughlin et al. (2022)	Nigeria, >555 children	IMCI data, symptoms	Sens 60%, spec 79%

© 2026 The Authors. Published by Bexford Publishing Ltd

Ojurongbe et al. (2023)	Nigeria, 200	Symptoms, environment	AUC 0.83 (0.63-1.00)
Bria et al. (2021)	Indonesia, 472	Symptoms, history	Accuracy ~85%
Khan et al. (2026)	Nigeria MIS, children	Household survey	Recall 83%
Mthethwa and Melesse (2025)	Nigeria MIS, 10,655	Household survey	AUC 0.80*
Ashaolu et al. (2025)	Nigeria IDP camps, 693	Household, caregiver	AUC 0.89
Loum et al. (2019)	Senegal, 15,523	Symptoms, rainfall	~92% (simulated test)
Shenoy et al. (2022)	India, 800	Clinical, laboratory	Accuracy 79-84%
Morang'a et al. (2020)	Ghana, 2,207	Haematology	AUC 0.87
Dharap and Raimbault (2020)	India, 1,420	Analyser flags	AUC 0.90 (external)
Balerdi-Sarasola et al. (2024)	Europe, 764 travellers	Clinical, laboratory	AUC 0.92-0.95
Ruban et al. (2026)	India, 1,455	Symptoms, signs	AUC 0.96
Mariki et al. (2022)	Tanzania, 2,556	Symptoms, demographics	Accuracy 82-95%
Nyengera et al. (2026)	Zimbabwe, 637	Symptoms, demographics	AUC 0.98*
Barraclough et al. (2025)	Kenya, 2,000	Symptoms, vitals	Accuracy 98.3%

Two studies applied SMOTE oversampling to the whole dataset before splitting, which leaks synthetic copies of test cases into training (Kapoor & Narayanan, 2023; Mbaye et al., 2021; Mthethwa & Melesse, 2025). Most studies used a single random split of a small sample, and few reported calibration, confidence intervals or decision thresholds. One overall rating was low risk of bias, 17 were moderate and 25 were high.

RQ1: Differential Diagnosis and Explanation

Six studies addressed malaria within a wider differential diagnosis, which is closer to the clinical question in febrile patients. Accuracy fell as the task became more realistic. A four-class model for malaria, typhoid, both or neither in Niger Delta data reached 72% accuracy (Attai et al., 2024). Among 800 patients with dengue, malaria, leptospirosis or scrub typhus, one-versus-rest logistic regression reached 79 to 84% accuracy while more complex learners reached 55 to 60% (Shenoy et al., 2022). In Senegal, separating malaria from arboviral infections relied on rebalanced and simulated evaluation data (Loum et al., 2019). A model separating Lassa fever from malaria reported 98.7% accuracy, with SMOTE correctly restricted to the training set (Esan et al., 2026). Explanation methods were added in several studies, including LIME and SHAP (Awe et al., 2025; Deepika & Pradeep Kumar, 2025) and explanations generated by language models (Amannah et al., 2025; Attai et al., 2025). These describe what a model has learned but cannot correct a model trained on circular or leaked labels.

RQ1: Free Text, Language Models and Feature Optimisation

Four studies used language. Clinical notes digitised by optical character recognition gave an F1 score of 0.76 for malaria (Ruban et al., 2021), and features extracted from published case reports gave a malaria

© 2026 The Authors. Published by Bexford Publishing Ltd

recall of at most 0.46 (Lee et al., 2021). A zero-shot evaluation of a general-purpose language model on case summaries of hospitalised febrile travellers reported a sensitivity of 100% and specificity of 94% for malaria (Yelin et al., 2025); the summaries included laboratory results, so the result is not a pre-test estimate. A symptom chatbot reported 98% detection but was trained on clinician-assigned labels (Oguntimilehin, Babalola, et al., 2024). Metaheuristic optimisation was rare: one study used an artificial immune algorithm to select symptoms for dengue, leptospirosis and malaria (Caicedo et al., 2012), and others used grid or random search for hyperparameters. No study combined structured clinical data with free text or speech for malaria.

RQ2: Speech and Clinical Text in Other Infections

The 103 RQ2 studies were dominated by COVID-19 (84 studies). The others covered tuberculosis, influenza and other respiratory infections, pneumonia, sepsis, urinary tract infection, invasive fungal disease, dengue and multi-disease symptom checkers. No RQ2 study targeted malaria. Table 4 groups the studies by input type. Sixty used speech or voice audio, often with cough or breath recordings; 13 combined voice or audio with clinical or symptom data; three converted spoken symptom descriptions to text or intents; 21 used clinical text alone; and six combined text with structured data. Thirty-nine used strong validation, but only 14 tested on external, temporal or cross-dataset data. Risk of bias was high in 51 studies, moderate in 48 and low in four (Figure 2).

Table 4: RQ2 Studies by Input Type and Validation

Input type	Sample	Strong validation	External or cross-dataset
Speech or voice audio	60	28	8
Speech or audio with clinical or symptom data	13	3	2
Spoken symptoms (speech to text or intent)	3	0	0
Clinical text only	21	6	3
Clinical text with structured data	6	2	1
Total	**103**	**39**	**14**

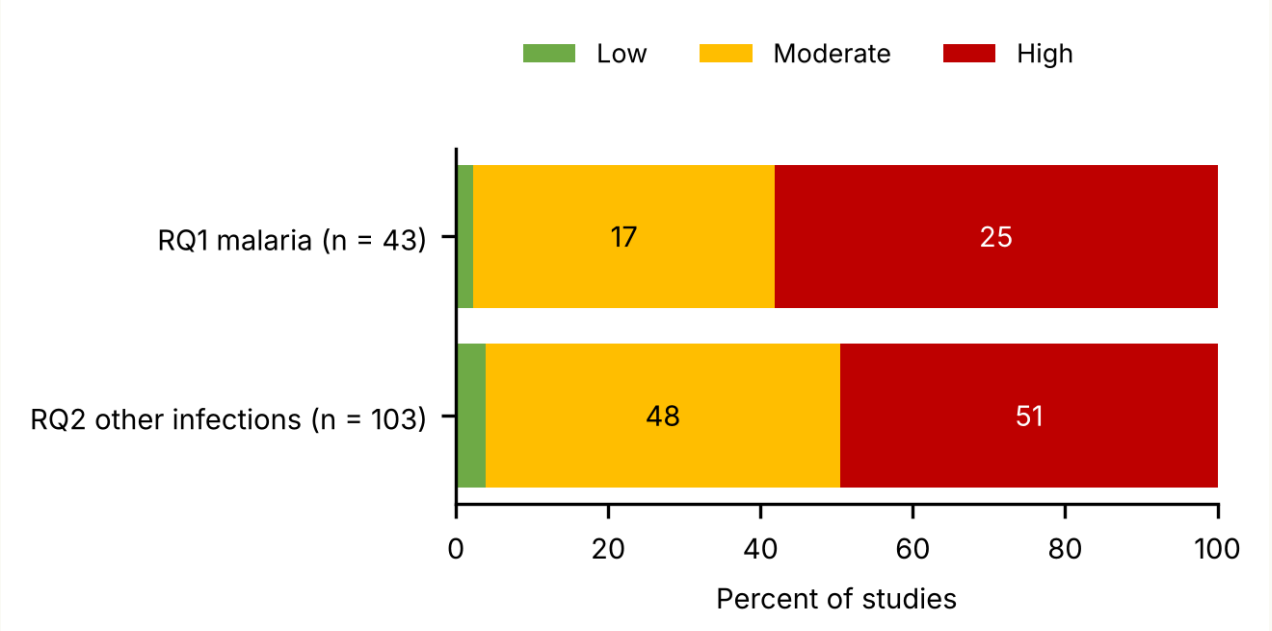


Figure 2: Overall risk of bias of included studies (adapted PROBAST); numbers in bars are study counts

© 2026 The Authors. Published by Bexford Publishing Ltd

RQ2: Generalisation of Speech Models

Within a single dataset, speech and voice models often reported accuracy above 90%, with some reports near 100% from sustained vowels or small clinical recordings (Costantini et al., 2022; Gidaye et al., 2025). Those results did not survive independent testing. When systems were trained on one challenge dataset and tested on another, performance fell sharply (Akman et al., 2022), in one comparison to chance level (Zhu et al., 2022). A Brazilian model for respiratory insufficiency that reached 96.5% accuracy on its first cohort fell below 39% on a later cohort that included other causes and hospital controls (Gauy et al., 2023). A system evaluated on four external datasets ranged from an AUC of 0.64 to 0.87 and degraded with recording device, background noise and symptom status (Grant et al., 2022). Speaker-dependent splits inflated results (Triantafyllopoulos et al., 2024), and when COVID-19 positive and negative speakers were balanced for symptoms, neither clinicians nor a baseline model could separate them (Hecker et al., 2021). The largest controlled analysis, not included here because it was not retrieved by our search, reached the same conclusion: audio classifiers did not outperform a simple symptom checker once confounders were controlled (Coppock et al., 2024). The most robust speech results were modest. A multi-site, multilingual vocal biomarker evaluated against PCR reached a sensitivity of 73.2% and specificity of 62.9% (Kaur, Larsen, et al., 2023), and participant-disjoint challenge benchmarks reported unweighted average recall of about 0.72 to 0.76 (Coppock et al., 2023; Koike et al., 2023).

Sequence and attention architectures appeared mainly in this group. A recurrent model of longitudinal voice, cough and breath recordings reached an AUROC of 0.79 with participant-level splits (Dang et al., 2022), contextual attention over self-supervised speech representations reached an AUC of 0.81 (Mallol-Ragolta et al., 2022), and transformer attention over three sound types reached 0.70 (Chang et al., 2021). In these studies the gain from the architecture was smaller than the losses caused by dataset shift.

RQ2: Fusion of Speech or Text with Clinical Data

Fusion methods are complex. For instance, symptoms and metadata combined with audio improved one challenge system AUC from 0.866 to 0.922 (Reiter & Pernkopf, 2025). Additionally, nine categories of audio and self-reported symptoms showed better performance compared with models using one audio category (Liu et al., 2026). Similarly, voice analysis was of comparable accuracy at 83 and 80%, respectively, to a six-symptom survey (Shimon et al., 2021). For instance, in Colombian tuberculosis records, two feature sets (structured clinical data and natural language processing features of clinical text) did not perform better when combined than just the structured clinical data (Orjuela-Cañón et al., 2025). However, for pneumonia diagnosis, imaging report text was important to the data, whereas symptom descriptions from clinical notes were not (DeLisle et al., 2013). In addition, an integrated model of urinalysis and symptom information from NLPs attained a urinary tract infection AUC of 0.93 (Iscoe et al., 2026). Most studies that obtained the near-perfect AUC using fusion, however, used different subjects' or sample-level splits in sample-level analysis or specified modalities (Jayachitra et al., 2021; Nguyen-Trong & Nguyen-Hoang, 2023).

RQ2: Clinical Text and Spoken Symptoms

Clinical-text models with laboratory-confirmed labels reported useful but not perfect discrimination, such as an AUC of 0.88 to 0.93 for influenza from emergency-department notes (López Pineda et al., 2015) and an F1 score of 0.56 for COVID-19 from hospital notes (Wang et al., 2023). Text models were vulnerable to drift: an NLP model for COVID-19 in primary-care notes lost performance sharply on later data (Hendrix et al., 2025), and a fine-tuned language model for tuberculosis from radiology reports fell to an F1 of 0.48 at an external hospital while a prompted general model held 0.86 (Li et al., 2025). Radiology reports and case reports often contain the diagnosis itself, which explains some very high accuracies (Khanday et al.,

© 2026 The Authors. Published by Bexford Publishing Ltd

2020; Qomariyah et al., 2022). Three studies classified diseases from spoken symptom descriptions; all used read or scripted speech and could only separate broad groups of conditions (Abdulmohsin, 2022; Abdulmohsin et al., 2022; Hossen et al., 2024).

Principal Findings

There are few studies on the subject of combining structured clinical data and free text or free verbal communication within the framework of malaria diagnosis, and none of them considers speech only as a means of diagnosing malaria. There are such studies that predict malaria before testing, and most of them are small, single-site studies focusing on symptoms. Accuracy depends greatly on the construction of diagnostic labels and data segmentation procedures. When parasitological testing was used to establish diagnostic labels and independent evaluations, symptom-based diagnostic models achieved 0.60-0.95 AUC and scored higher only when laboratory results were included. Such variability fits with current clinical understanding of overlap of symptoms between malaria and other febrile diseases.

Conversely, speech studies of other infectious diseases serve as a caveat rather than a recipe to follow. While aspects of voice provide reflection of diseases of the respiratory tract, uncontrolled factors of speaker differences, device differences, language differences, and recruitment methods used to draw cases and controls will conflate and exaggerate performance indicators based on data. When controlling for these factors, speech features did not add much to the symptomatology (Coppock et al., 2024; Hecker et al., 2021). However, in some cases, data fusion was successful and in others, not, even after changing the structure of data.

Language, Setting and Language Models

In Nigeria, for example, patients collect their health problems in different languages spoken between them (English, Nigerian Pidgin, Yoruba, Hausa, Igbo, etc.). Speech models built for a certain language and in a certain recording environment do not perform well if they are used in another one (Gauy et al., 2023; Grant et al., 2022). We should therefore analyse malaria-related speech models by patient language and device type. General language models performed well on carefully selected case summaries (Yelin et al., 2025) and outperformed specific fine-tuned language models at an external hospital (Li et al., 2025), which suggests their potential for text input. Nevertheless, their use for triage pre-testing should be restricted to information known prior to testing and locally validated.

Implications for Pre-Test Malaria Models

Table 5 converts these results to prerequisites for any future multimodal model for pre-test malaria triage. Parasitological labels, patient-specific and speaker-neutral divisions, controls from febrile patients with negative results and not healthy volunteers, external validation performed at another location and ablations that capture each input individually and collectively are essential. All reporting should follow TRIPOD+AI (Collins et al., 2024), encompassing calibration and decision-curve or threshold analysis, as the effectiveness of a triage tool should be based on the changes it makes to the testing and treatment decision rather than precision. Free text should be restricted to the period prior to test results to avoid loss of information in radiology and case report analysis.

Table 5: Requirements for Pre-Test Multimodal Malaria Models

Requirement	Evidence from this review
Label every case with microscopy, RDT or PCR	Clinician or dataset labels in 27 of 43 RQ1 studies; inflated accuracy

© 2026 The Authors. Published by Bexford Publishing Ltd

Exclude test results and diagnoses from inputs, including text written after the test	100% accuracy when the result or codes were inputs; leakage in report text
Split by patient, and by speaker for voice; resample only training folds	SMOTE before splitting; speaker-dependent splits inflated results
Use febrile, test-negative controls and balance symptoms	Healthy-volunteer controls and symptom imbalance drove speech results
Validate externally on another site or period	Only 2 of 43 RQ1 and 14 of 103 RQ2 studies; large drops when done
Report each input alone and fused, against a symptom-only baseline	Fusion helped in some studies and not in others
Report calibration, thresholds and confidence intervals (TRIPOD+AI)	Rarely reported in either group

Limitations

The search covered two major databases (Scopus and PubMed), so studies indexed only in IEEE Xplore, the ACM Digital Library or Web of Science were not included. Only English-language records were included. Twenty-two reports could not be retrieved. Screening, extraction and verification were done by the authors. Risk-of-bias ratings rely on what studies reported, and many reported too little to judge. The RQ2 evidence is dominated by COVID-19, whose acoustic effects may differ from those of malaria.

V. CONCLUSION

Machine learning applications for diagnosis of malaria prior to testing are still under development. However, most existing models are not reliable for clinical use due to lack of labelling or validation to substantiate their accuracy. Speech and clinical text as potential diagnosis variables for malaria are still not addressed, even though evidence from other infectious diseases shows that their effectiveness can only be confirmed using symptom-matched controls, patient-based analyses, speaker-independent divisions, and external validation. Future studies should focus on multimodal pre-test models based on prospectively collected data from parasitologically confirmed febrile cohorts in endemic regions to comply with the TRIPOD+AI reporting standards.

References

Abdulmohsin, H. A. (2022). Automatic health speech prediction system using support vector machine. In *Proceedings of International Conference on Computing and Communication Networks* (pp. 165–175). Springer. https://doi.org/10.1007/978-981-19-0604-6_15

Abdulmohsin, H. A., Al-Khateeb, B., Hasan, S. S., & Dwivedi, R. (2022). Automatic illness prediction system through speech. *Computers and Electrical Engineering, 102*, Article 108224. https://doi.org/10.1016/j.compeleceng.2022.108224

Achary, R., Shelke, C. J., & Lekhya, A. (2025). A GAN-enhanced multimodal diagnostic framework utilizing an ensemble of BiLSTM, BiGRU, and RNN models for malaria and dengue detection. *Procedia Computer Science, 252*, 381–393. https://doi.org/10.1016/j.procs.2024.12.039

Acosta, J. N., Falcone, G. J., Rajpurkar, P., & Topol, E. J. (2022). Multimodal biomedical AI. *Nature Medicine, 28*(9), 1773–1784. https://doi.org/10.1038/s41591-022-01981-2

© 2026 The Authors. Published by Bexford Publishing Ltd

Ahamed, M. F., Nahiduzzaman, M., Mahmud, G., Shafi, F. B., Ayari, M. A., Khandakar, A., Abdullah-Al-Wadud, M., & Islam, S. M. R. (2025). Improving malaria diagnosis through interpretable customized CNNs architectures. *Scientific Reports, 15*(1), Article 6484. https://doi.org/10.1038/s41598-025-90851-1

Akman, A., Coppock, H., Gaskell, A., Tzirakis, P., Jones, L., & Schuller, B. W. (2022). Evaluating the COVID-19 identification ResNet (CIdeR) on the INTERSPEECH COVID-19 from audio challenges. *Frontiers in Digital Health, 4*, Article 789980. https://doi.org/10.3389/fdgth.2022.789980

AlSaad, R., Abd-alrazaq, A., Boughorbel, S., Ahmed, A., Renault, M.-A., Damseh, R., & Sheikh, J. (2024). Multimodal large language models in health care: Applications, challenges, and future outlook. *Journal of Medical Internet Research, 26*, Article e59505. https://doi.org/10.2196/59505

Amannah, C., Attai, K. F., & Uzoka, F.-M. (2025). A data-driven intelligent methodology for developing explainable diagnostic model for febrile diseases. *Algorithms, 18*(4), Article 190. https://doi.org/10.3390/a18040190

Amrevuawho, O. F., Adebiyi, M. O., Adebiyi, A. A., Ogbonnia, E. O., Alfa, P. E., & Eden, S. K. (2025). Unimodal to multimodal machine learning in malaria diagnosis: Challenges and opportunities. *NIPES Journal of Science and Technology Research, 7*(1), 1068–1074. https://doi.org/10.37933/nipes/7.4.2025.SI122

Anjorin, E. T., Olulaja, O. N., Osoba, M. E., Oyadiran, O. T., Ogunsanya, A. O., Akinade, O. N., & Inuojo, J. M. (2023). Overtreatment of malaria in the Nigerian healthcare setting: Prescription practice, rationale and consequences. *Pan African Medical Journal, 45*, Article 111. https://doi.org/10.11604/pamj.2023.45.111.31780

Ashaolu, J. O., Akanji, T. S., Ayansola, V. I., Olawale-Succes, O. O., Sunday, A. J., & Some, S. Y. M. (2025). From risk factors to predictive modelling: Applying machine learning to childhood malaria surveillance in resource-limited settings. *BMC Infectious Diseases, 25*(1), Article 1693. https://doi.org/10.1186/s12879-025-12116-6

Atoyebi, T. O., Olanrewaju, R. F., Blamah, N. V., & Uwazie, E. C. (2024). Comparison of multinomial naive Bayes (MNB), Gaussian naive Bayes (GNB) and random forest (RF) algorithm in malaria disease diagnosis. In *2024 International Conference on Science, Engineering and Business for Driving Sustainable Development Goals (SEB4SDG)* (pp. 1–6). IEEE. https://doi.org/10.1109/seb4sdg60871.2024.10630308

Attai, K., Ekpenyong, M., Amannah, C., Asuquo, D., Ajuga, P., Obot, O., Johnson, E., John, A., Maduka, O., Akwaowo, C., & Uzoka, F.-M. (2024). Enhancing the interpretability of malaria and typhoid diagnosis with explainable AI and large language models. *Tropical Medicine and Infectious Disease, 9*(9), Article 216. https://doi.org/10.3390/tropicalmed9090216

Attai, K. F., Amannah, C., Ekpenyong, M., Asuquo, D. E., Akputu, O. K., Obot, O. U., Ajuga, P. C., Obi, J. C., Maduka, O., Akwaowo, C., & Uzoka, F.-M. (2025). Developing an explainable artificial intelligence system for the mobile-based diagnosis of febrile diseases using random forest, LIME, and GPT. *Healthcare Informatics Research, 31*(2), 125–135. https://doi.org/10.4258/hir.2025.31.2.125

Awe, O. O., Mwangi, P. N., Goudoungou, S. K., Esho, R. V., & Oyejide, O. S. (2025). Explainable AI for enhanced accuracy in malaria diagnosis using ensemble machine learning models. *BMC Medical Informatics and Decision Making, 25*(1), Article 162. https://doi.org/10.1186/s12911-025-02874-3

Ayalew, A. M., Admass, W. S., Abuhayi, B. M., Negashe, G. S., & Bezabh, Y. A. (2024). Smart malaria

© 2026 The Authors. Published by Bexford Publishing Ltd

classification: A novel machine learning algorithms for early malaria monitoring and detecting using IoT-based healthcare environment. *Sensing and Imaging, 25*(1), Article 55. https://doi.org/10.1007/s11220-024-00503-3

Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. In *Proceedings of the 3rd International Conference on Learning Representations*. https://doi.org/10.48550/arXiv.1409.0473

Balerdi-Sarasola, L., Fleitas, P., Bottieau, E., Genton, B., Petrone, P., Muñoz, J., & Camprubí-Ferrer, D. (2024). MALrisk: A machine-learning-based tool to predict imported malaria in returned travellers with fever. *Journal of Travel Medicine, 31*(8), Article taae054. https://doi.org/10.1093/jtm/taae054

Baltrusaitis, T., Ahuja, C., & Morency, L.-P. (2019). Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 41*(2), 423–443. https://doi.org/10.1109/TPAMI.2018.2798607

Barraclough, P. A., Were, C. M., Mwangakala, H., Anderson, P., Ohanya, D. O., Agola, H., & Nandi, P. (2025). Intelligent diagnostic model for early malaria symptoms. *International Journal of Advanced Computer Science and Applications, 16*(12), 72–82. https://doi.org/10.14569/ijacsa.2025.0161207

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research, 13*, 281–305. https://jmlr.org/papers/v13/bergstra12a.html

Bischl, B., Binder, M., Lang, M., Pielok, T., Richter, J., Coors, S., Thomas, J., Ullmann, T., Becker, M., Boulesteix, A.-L., Deng, D., & Lindauer, M. (2023). Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges. *WIREs Data Mining and Knowledge Discovery, 13*(2), Article e1484. https://doi.org/10.1002/widm.1484

Boit, S., & Patil, R. (2024). An efficient deep learning approach for malaria parasite detection in microscopic images. *Diagnostics, 14*(23), Article 2738. https://doi.org/10.3390/diagnostics14232738

Bria, Y. P., Yeh, C.-H., & Bedingfield, S. (2021). Significant symptoms and nonsymptom-related factors for malaria diagnosis in endemic regions of Indonesia. *International Journal of Infectious Diseases, 103*, 194–200. https://doi.org/10.1016/j.ijid.2020.11.177

Caicedo, W., Quintana, M., & Pinzón, H. (2012). Differential diagnosis of hemorrhagic fevers using ARTMAP. In *Advances in Artificial Intelligence – IBERAMIA 2012* (pp. 221–230). Springer. https://doi.org/10.1007/978-3-642-34654-5_23

Chang, Y., Ren, Z., & Schuller, B. W. (2021). Transformer-based CNNs: Mining temporal context information for multi-sound COVID-19 diagnosis. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (pp. 2335–2338). IEEE. https://doi.org/10.1109/embc46164.2021.9629552

Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing* (pp. 1724–1734). Association for Computational Linguistics. https://doi.org/10.3115/v1/D14-1179

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37–46. https://doi.org/10.1177/001316446002000104

Collins, G. S., Moons, K. G. M., Dhiman, P., Riley, R. D., Beam, A. L., Van Calster, B., Ghassemi, M., Liu, X., Reitsma, J. B., van Smeden, M., Boulesteix, A.-L., Camaradou, J. C., Celi, L. A., Denaxas,

© 2026 The Authors. Published by Bexford Publishing Ltd

S., Denniston, A. K., Glocker, B., Golub, R. M., Harvey, H., Heinze, G., . . . Logullo, P. (2024). TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ, 385*, Article e078378. https://doi.org/10.1136/bmj-2023-078378

Coppock, H., Akman, A., Bergler, C., Gerczuk, M., Brown, C., Chauhan, J., Grammenos, A., Hasthanasombat, A., Spathis, D., Xia, T., Cicuta, P., Han, J., Amiriparian, S., Baird, A., Stappen, L., Ottl, S., Tzirakis, P., Batliner, A., Mascolo, C., & Schuller, B. W. (2023). A summary of the ComParE COVID-19 challenges. *Frontiers in Digital Health, 5*, Article 1058163. https://doi.org/10.3389/fdgth.2023.1058163

Coppock, H., Nicholson, G., Kiskin, I., Koutra, V., Baker, K., Budd, J., Payne, R., Karoune, E., Hurley, D., Titcomb, A., Egglestone, S., Tendero Cañadas, A., Butler, L., Jersakova, R., Mellor, J., Patel, S., Thornley, T., Diggle, P., Richardson, S., . . . Holmes, C. (2024). Audio-based AI classifiers show no evidence of improved COVID-19 screening over simple symptoms checkers. *Nature Machine Intelligence, 6*(2), 229–242. https://doi.org/10.1038/s42256-023-00773-8

Costantini, G., Cesarini, V., Robotti, C., Benazzo, M., Pietrantonio, F., Di Girolamo, S., Pisani, A., Canzi, P., Mauramati, S., Bertino, G., Cassaniti, I., Baldanti, F., & Saggio, G. (2022). Deep learning and machine learning-based voice analysis for the detection of COVID-19: A proposal and comparison of architectures. *Knowledge-Based Systems, 253*, Article 109539. https://doi.org/10.1016/j.knosys.2022.109539

Dang, T., Han, J., Xia, T., Spathis, D., Bondareva, E., Siegele-Brown, C., Chauhan, J., Grammenos, A., Hasthanasombat, A., Floto, R. A., Cicuta, P., & Mascolo, C. (2022). Exploring longitudinal cough, breath, and voice data for COVID-19 progression prediction via sequential deep learning: Model development and validation. *Journal of Medical Internet Research, 24*(6), Article e37004. https://doi.org/10.2196/37004

Deepika, R., & Pradeep Kumar, T. S. (2025). An interpretable dual-level lightweight framework for malaria diagnosis for enhanced remote diagnosis in resource-constrained settings. *Systems Science and Control Engineering, 13*(1), Article 2579988. https://doi.org/10.1080/21642583.2025.2579988

DeLisle, S., Kim, B., Deepak, J., Siddiqui, T., Gundlapalli, A., Samore, M., & D'Avolio, L. (2013). Using the electronic medical record to identify community-acquired pneumonia: Toward a replicable automated strategy. *PLoS ONE, 8*(8), Article e70944. https://doi.org/10.1371/journal.pone.0070944

Dharap, P., & Raimbault, S. (2020). Performance evaluation of machine learning-based infectious screening flags on the HORIBA Medical Yumizen H550 haematology analyzer for vivax malaria and dengue fever. *Malaria Journal, 19*(1), Article 429. https://doi.org/10.1186/s12936-020-03502-3

Eguia, H., Sánchez-Bocanegra, C. L., Vinciarelli, F., Alvarez-Lopez, F., & Saigí-Rubió, F. (2024). Clinical decision support and natural language processing in medicine: Systematic literature review. *Journal of Medical Internet Research, 26*, Article e55315. https://doi.org/10.2196/55315

Esan, A., Adejo, G., Okomba, N., Soladoye, A. A., Aderinto, N., & Olawade, D. B. (2026). AI-driven diagnosis of Lassa fever: Evidence from Nigerian clinical records. *Computational Biology and Chemistry, 120*, Article 108627. https://doi.org/10.1016/j.compbiolchem.2025.108627

Gauy, M. M., Berti, L. C., Cândido, A., Jr., Neto, A. C., Goldman, A., Levin, A. S. S., Martins, M., de Medeiros, B. R., Queiroz, M., Sabino, E. C., Svartman, F. R. F., & Finger, M. (2023). Discriminant audio properties in deep learning based respiratory insufficiency detection in Brazilian Portuguese. In *Artificial Intelligence in Medicine* (pp. 271–275). Springer. https://doi.org/10.1007/978-3-031-34344-5_32

© 2026 The Authors. Published by Bexford Publishing Ltd

Gidaye, G., Barage, A., Dighe, N., Ezzine, K., Turkar, V., & Nagare, G. (2025). Speech signals as biomarkers: Using glottal features for non-invasive COVID-19 testing. *International Journal of Biomedical Engineering and Technology, 47*(1), 65–85. https://doi.org/10.1504/ijbet.2025.143755

Grant, D., McLane, I., Rennoll, V., & West, J. (2022). Considerations and challenges for real-world deployment of an acoustic-based COVID-19 screening system. *Sensors, 22*(23), Article 9530. https://doi.org/10.3390/s22239530

Hecker, P., Pokorny, F. B., Bartl-Pokorny, K. D., Reichel, U., Ren, Z., Hantke, S., Eyben, F., Schuller, D. M., Arnrich, B., & Schuller, B. W. (2021). Speaking corona? Human and machine recognition of COVID-19 from voice. In *Proceedings of Interspeech 2021* (pp. 1029–1033). International Speech Communication Association. https://doi.org/10.21437/interspeech.2021-1771

Hendrix, N., Parikh, R. V., Taskier, M., Walter, G., Phillips, R. L., & Rehkopf, D. H. (2025). Natural language processing improves reliable identification of COVID-19 compared to diagnostic codes alone. *American Journal of Epidemiology, 194*(11), 3348–3354. https://doi.org/10.1093/aje/kwaf162

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hossen, J., Islam, M. R., Chowdhury, A., Ukti, I. J., & Islam, M. M. (2024). DocBot: A system for disease detection and specialized doctor recommendation using patient's speech of symptoms. In *Emerging Technologies in Computing* (pp. 71–85). Springer. https://doi.org/10.1007/978-3-031-50215-6_6

Hoyos, K., & Hoyos, W. (2024). Supporting malaria diagnosis using deep learning and data augmentation. *Diagnostics, 14*(7), Article 690. https://doi.org/10.3390/diagnostics14070690

Hu, S., Oppong, A., Mogo, E., Collins, C., Occhini, G., Barford, A., & Korhonen, A. (2025). Natural language processing technologies for public health in Africa: Scoping review. *Journal of Medical Internet Research, 27*, Article e68720. https://doi.org/10.2196/68720

Iscoe, M., Li, H., Xue, H., Socrates, V., Gilson, A., Huang, T., & Taylor, R. A. (2026). Evaluating the potential impact of AI on urinary tract infection diagnosis in the emergency department across demographic groups: Retrospective cohort study. *JMIR AI, 5*, Article e91148. https://doi.org/10.2196/91148

Jayachitra, V. P., Nivetha, S., Nivetha, R., & Harini, R. (2021). A cognitive IoT-based framework for effective diagnosis of COVID-19 using multimodal data. *Biomedical Signal Processing and Control, 70*, Article 102960. https://doi.org/10.1016/j.bspc.2021.102960

Kapoor, S., & Narayanan, A. (2023). Leakage and the reproducibility crisis in machine-learning-based science. *Patterns, 4*(9), Article 100804. https://doi.org/10.1016/j.patter.2023.100804

Kaur, S., Kumar, Y., Koul, A., & Kumar Kamboj, S. (2023). A systematic review on metaheuristic optimization techniques for feature selections in disease diagnosis: Open issues and challenges. *Archives of Computational Methods in Engineering, 30*(3), 1863–1895. https://doi.org/10.1007/s11831-022-09853-1

Kaur, S., Larsen, E., Harper, J., Purandare, B., Uluer, A., Hasdianda, M. A., Umale, N. A., Killeen, J., Castillo, E., & Jariwala, S. (2023). Development and validation of a respiratory-responsive vocal biomarker–based tool for generalizable detection of respiratory impairment: Independent case-control studies in multiple respiratory conditions including asthma, chronic obstructive pulmonary disease, and COVID-19. *Journal of Medical Internet Research, 25*, Article e44410. https://doi.org/10.2196/44410

Khan, A. R., Mahbub, N., Aziz, R. A., Siddique, M. H., & Sayeem, Z. I. (2026). Early malaria risk

© 2026 The Authors. Published by Bexford Publishing Ltd

screening in Nigerian minors using AutoML and cluster-based analysis of non-clinical survey data. *PLOS Digital Health, 5*(9), Article e0001736. https://doi.org/10.1371/journal.pdig.0001736

Khanday, A. M. U. D., Rabani, S. T., Khan, Q. R., Rouf, N., & Mohi Ud Din, M. (2020). Machine learning based approaches for detecting COVID-19 using clinical text data. *International Journal of Information Technology, 12*(3), 731–739. https://doi.org/10.1007/s41870-020-00495-9

Kline, A., Wang, H., Li, Y., Dennis, S., Hutch, M., Xu, Z., Wang, F., Cheng, F., & Luo, Y. (2022). Multimodal machine learning in precision health: A scoping review. *npj Digital Medicine, 5*(1), Article 171. https://doi.org/10.1038/s41746-022-00712-8

Koike, T., Wang, Z., Qian, K., Hu, B., Schuller, B. W., & Yamamoto, Y. (2023). An investigation on data augmentation and multiple instance learning for diagnosis of COVID-19 from speech and cough sound. In *2023 International Conference on Consumer Electronics – Taiwan (ICCE-Taiwan)* (pp. 783–784). IEEE. https://doi.org/10.1109/icce-taiwan58799.2023.10226678

Lee, Y. W., Choi, J. W., & Shin, E.-H. (2021). Machine learning model for predicting malaria using clinical information. *Computers in Biology and Medicine, 129*, Article 104151. https://doi.org/10.1016/j.compbiomed.2020.104151

Li, R., Mao, S., Zhu, C., Yang, Y., Tan, C., Li, L., Mu, X., Liu, H., & Yang, Y. (2025). Enhancing pulmonary disease prediction using large language models with feature summarization and hybrid retrieval-augmented generation: Multicenter methodological study based on radiology report. *Journal of Medical Internet Research, 27*, Article e72638. https://doi.org/10.2196/72638

Liu, Z., De Bock, K. W., Zhang, L., Wang, J., & Liang, L. (2026). A heterogeneous ensemble learning framework for detecting acute viral respiratory infections via multisource data fusion. *Annals of Operations Research, 362*(1–3), 521–551. https://doi.org/10.1007/s10479-026-07264-0

Loum, M. A., Poursat, M.-A., Sow, A., Sall, A. A., Loucoubar, C., & Gassiat, E. (2019). Multinomial logistic model for coinfection diagnosis between arbovirus and malaria in Kedougou. *International Journal of Biostatistics, 15*(2), Article 20170015. https://doi.org/10.1515/ijb-2017-0015

López Pineda, A., Ye, Y., Visweswaran, S., Cooper, G. F., Wagner, M. M., & Tsui, F. R. (2015). Comparison of machine learning classifiers for influenza detection from emergency department free-text reports. *Journal of Biomedical Informatics, 58*, 60–69. https://doi.org/10.1016/j.jbi.2015.08.019

Mallol-Ragolta, A., Liu, S., & Schuller, B. (2022). COVID-19 detection exploiting self-supervised learning representations of respiratory sounds. In *2022 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)* (pp. 1–4). IEEE. https://doi.org/10.1109/bhi56158.2022.9926967

Mariki, M., Mkoba, E., & Mduma, N. (2022). Combining clinical symptoms and patient features for malaria diagnosis: Machine learning approach. *Applied Artificial Intelligence, 36*(1), Article 2031826. https://doi.org/10.1080/08839514.2022.2031826

Mbaye, O., Ba, M. L., & Sy, A. (2021). On the efficiency of machine learning models in malaria prediction. In *Public Health and Informatics: Proceedings of MIE 2021* (pp. 437–441). IOS Press. https://doi.org/10.3233/shti210196

McLaughlin, M., Pellé, K. G., Scarpino, S. V., Giwa, A., Mount-Finette, E., Haidar, N., Adamu, F., Ravi, N., Thompson, A., Heath, B., Dittrich, S., & Finette, B. (2022). Development and validation of manually modified and supervised machine learning clinical assessment algorithms for malaria in Nigerian children. *Frontiers in Artificial Intelligence, 4*, Article 554017. https://doi.org/10.3389/frai.2021.554017

© 2026 The Authors. Published by Bexford Publishing Ltd

Morang'a, C. M., Amenga-Etego, L., Bah, S. Y., Appiah, V., Amuzu, D. S. Y., Amoako, N., Abugri, J., Oduro, A. R., Cunnington, A. J., Awandare, G. A., & Otto, T. D. (2020). Machine learning approaches classify clinical malaria outcomes based on haematological parameters. *BMC Medicine, 18*(1), Article 375. https://doi.org/10.1186/s12916-020-01823-3

Mthethwa, S. M., & Melesse, S. F. (2025). A machine learning approach to the prediction of malaria in under-five children: Analysis of the 2021 Nigerian malaria indicator survey. *The Open Public Health Journal, 18*(1), Article e18749445396163. https://doi.org/10.2174/0118749445396163250604103305

Muguli, A., Pinto, L., Nirmala, R., Sharma, N., Krishnan, P., Ghosh, P. K., Kumar, R., Bhat, S., Chetupalli, S. R., Ganapathy, S., Ramoji, S., & Nanda, V. (2021). DiCOVA challenge: Dataset, task, and baseline system for COVID-19 diagnosis using acoustics. In *Proceedings of Interspeech 2021* (pp. 901–905). International Speech Communication Association. https://doi.org/10.21437/interspeech.2021-74

Nguyen-Trong, K., & Nguyen-Hoang, K. (2023). Multi-modal approach for COVID-19 detection using coughs and self-reported symptoms. *Journal of Intelligent and Fuzzy Systems, 44*(3), 3501–3513. https://doi.org/10.3233/jifs-222863

Nyengera, P., Takawira, H. T., & Mlambo, F. F. (2026). Ensemble machine learning for malaria diagnosis in resource-limited settings using clinical and demographic features. *Infectious Disease Reports, 18*(4), Article 72. https://doi.org/10.3390/idr18040072

Obeagu, E. I., & Ogenyi, F. C. (2025). Artificial intelligence in African malaria control programs: Opportunities and risks. *Annals of Medicine and Surgery, 87*(12), 8664–8670. https://doi.org/10.1097/MS9.0000000000004259

Oguntimilehin, A., Babalola, G. O., Akinduyite, C. O., Obamiyi, S. E., Okebule, T., Oyelami, F. H., Abiola, O. B., Badeji-Ajisafe, B., & Abang, E. (2024). A machine learning driven malaria fever medical consultation chatbot. In *2024 International Conference on Science, Engineering and Business for Driving Sustainable Development Goals (SEB4SDG)* (pp. 1–6). IEEE. https://doi.org/10.1109/seb4sdg60871.2024.10629694

Oguntimilehin, A., Babalola, G. O., & Olatunji, K. A. (2019). A clinical diagnostic model based on supervised learning. *International Journal of Advanced Trends in Computer Science and Engineering, 8*(3), 949–953. https://doi.org/10.30534/ijatcse/2019/94832019

Oguntimilehin, A., Berida, O. T., Abiola, O. B., Badeji-Ajisafe, B., Okebule, T., Akinduyite, C. O., Obamiyi, S. E., Babalola, G. O., & Sanya, O. (2024). A multilayer perceptron based mobile diagnosis system for malaria fever. In *2024 International Conference on Science, Engineering and Business for Driving Sustainable Development Goals (SEB4SDG)* (pp. 1–8). IEEE. https://doi.org/10.1109/seb4sdg60871.2024.10629733

Ojurongbe, T. A., Afolabi, H. A., Bashiru, K. A., Sule, W. F., Akinde, S. B., Ojurongbe, O., & Adegoke, N. A. (2023). Prediction of malaria positivity using patients' demographic and environmental features and clinical symptoms to complement parasitological confirmation before treatment. *Tropical Diseases, Travel Medicine and Vaccines, 9*(1), Article 24. https://doi.org/10.1186/s40794-023-00208-7

Okagbue, H. I., Oguntunde, P. E., Obasi, E. C. M., Adamu, P. I., & Opanuga, A. A. (2021). Diagnosing malaria from some symptoms: A machine learning approach and public health implications. *Health and Technology, 11*(1), 23–37. https://doi.org/10.1007/s12553-020-00488-5

Orjuela-Cañón, A. D., Romero-Gómez, A. F., Jutinico, A. L., Awad, C. E., Vergara, E., & Palencia, M. A.

© 2026 The Authors. Published by Bexford Publishing Ltd

(2025). Data fusion of medical records and clinical data to enhance tuberculosis diagnosis in resource-limited settings. *Applied Sciences, 15*(10), Article 5423. https://doi.org/10.3390/app15105423

Padmakala, S. (2026). Predictive intelligence to early intervention from malaria: Comparative study between quantum and classical machine learning paradigm. In *2026 IEEE 2nd International Conference on Quantum Photonics, Artificial Intelligence & Networking (QPAIN)* (pp. 1–6). IEEE. https://doi.org/10.1109/qpain69676.2026.11545930

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., . . . Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, Article n71. https://doi.org/10.1136/bmj.n71

Qomariyah, N. N., Araminta, A. S., Reynaldi, R., Senjaya, M., Asri, S. D. A., & Kazakov, D. (2022). NLP text classification for COVID-19 automatic detection from radiology report in Indonesian language. In *2022 5th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)* (pp. 565–569). IEEE. https://doi.org/10.1109/isriti56927.2022.10053077

Reiter, M., & Pernkopf, F. (2025). Acoustic COVID-19 detection using multiple instance learning. *IEEE Journal of Biomedical and Health Informatics, 29*(1), 620–630. https://doi.org/10.1109/jbhi.2024.3474975

Ruban, S., Naresh, A., & Rai, S. (2021). A noninvasive model to detect malaria based on symptoms using machine learning. *Advances in Parallel Computing, 39*, 23–30. https://doi.org/10.3233/apc210119

Ruban, S., Vinay, V., Shanthini, H., Rai, S., & Subramani, S. (2026). Leveraging machine learning for symptom-based malaria diagnosis: A predictive model approach. *Neural Computing and Applications, 38*(13), Article 566. https://doi.org/10.1007/s00521-026-12305-7

Sharma, N., Krishnan, P., Kumar, R., Ramoji, S., Chetupalli, S. R., Nirmala, R., Ghosh, P. K., & Ganapathy, S. (2020). Coswara—A database of breathing, cough, and voice sounds for COVID-19 diagnosis. In *Proceedings of Interspeech 2020* (pp. 4811–4815). International Speech Communication Association. https://doi.org/10.21437/interspeech.2020-2768

Shenoy, S., Rajan, A. K., Rashid, M., Chandran, V. P., Poojari, P. G., Kunhikatta, V., Acharya, D., Nair, S., Varma, M., & Thunga, G. (2022). Artificial intelligence in differentiating tropical infections: A step ahead. *PLoS Neglected Tropical Diseases, 16*(6), Article e0010455. https://doi.org/10.1371/journal.pntd.0010455

Shickel, B., Tighe, P. J., Bihorac, A., & Rashidi, P. (2018). Deep EHR: A survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. *IEEE Journal of Biomedical and Health Informatics, 22*(5), 1589–1604. https://doi.org/10.1109/JBHI.2017.2767063

Shimon, C., Shafat, G., Dangoor, I., & Ben-Shitrit, A. (2021). Artificial intelligence enabled preliminary diagnosis for COVID-19 from voice cues and questionnaires. *Journal of the Acoustical Society of America, 149*(2), 1120–1124. https://doi.org/10.1121/10.0003434

Stahlschmidt, S. R., Ulfenborg, B., & Synnergren, J. (2022). Multimodal deep learning for biomedical data fusion: A review. *Briefings in Bioinformatics, 23*(2), Article bbab569. https://doi.org/10.1093/bib/bbab569

Taneja, S. B., Douglas, G. P., Cooper, G. F., Michaels, M. G., Druzdzel, M. J., & Visweswaran, S. (2021). Bayesian network models with decision tree analysis for management of childhood malaria in

© 2026 The Authors. Published by Bexford Publishing Ltd

Malawi. *BMC Medical Informatics and Decision Making, 21*(1), Article 158. https://doi.org/10.1186/s12911-021-01514-w

Triantafyllopoulos, A., Semertzidou, A., Song, M., Pokorny, F. B., & Schuller, B. W. (2024). Introducing the COVID-19 YouTube (COVYT) speech dataset featuring the same speakers with and without infection. *Biomedical Signal Processing and Control, 88*, Article 105642. https://doi.org/10.1016/j.bspc.2023.105642

Wang, L., Zipursky, A. R., Geva, A., McMurry, A. J., Mandl, K. D., & Miller, T. A. (2023). A computable case definition for patients with SARS-CoV2 testing that occurred outside the hospital. *JAMIA Open, 6*(3), Article ooad047. https://doi.org/10.1093/jamiaopen/ooad047

Wolff, R. F., Moons, K. G. M., Riley, R. D., Whiting, P. F., Westwood, M., Collins, G. S., Reitsma, J. B., Kleijnen, J., & Mallett, S. (2019). PROBAST: A tool to assess the risk of bias and applicability of prediction model studies. *Annals of Internal Medicine, 170*(1), 51–58. https://doi.org/10.7326/M18-1376

World Health Organization. (2025a). *WHO guidelines for malaria* (13 August 2025 version). https://www.who.int/publications/i/item/guidelines-for-malaria

World Health Organization. (2025b). *World malaria report 2025*. https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2025

Yelin, D., Shirin, N., Harris, I., Peretz, Y., Yahav, D., Schwartz, E., Leshem, E., & Margalit, I. (2025). Performance of ChatGPT-4o in the diagnostic workup of fever among returning travellers requiring hospitalization: A validation study. *Journal of Travel Medicine, 32*(4), Article taaf005. https://doi.org/10.1093/jtm/taaf005

Zhu, Y., Mariakakis, A., De Lara, E., & Falk, T. H. (2022). How generalizable and interpretable are speech-based COVID-19 detection systems?: A comparative analysis and new system proposal. In *2022 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)* (pp. 1–5). IEEE. https://doi.org/10.1109/bhi56158.2022.9926950

© 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the **Creative Commons Attribution (CC BY 4.0) licence** (https://creativecommons.org/licenses/by/4.0/)