

Random Forest Algorithm Applications in Studying Children Under-five: Systematic Literature Review (SLR)

Maroning Useng¹, Sahidan Abdulmana¹, and Anas Tawalbeh^{1*}

¹Faculty of Science and Technology, Fatoni University, Pattani 94160, Thailand

* Corresponding Author, E-mail: anas.tawalbeh@ftu.ac.th DOI: 10.14416/JASET.KMUTNB. 2026.01.003

Received 22 December 2025; **Revised** 28 April 2026; **Accepted** 29 April 2026

ABSTRACT

Effective monitoring of health and development in children under five is essential for timely detection, informed clinical decision-making, and cost-effective public health action. Recent advances in machine learning have expanded the potential of under-five monitoring systems, particularly for early screening, population-level risk prediction, and determinant-oriented surveillance. This paper presents a systematic literature review of Random Forest and closely related modelling practices in under-five monitoring, with attention to both opportunities and methodological constraints. The review identified 15 eligible studies, covering applications in respiratory infection monitoring, malaria risk prediction, nutrition and growth surveillance, and developmental assessment, drawing on national survey microdata, clinical records, and imaging or audio modalities. The review concludes by outlining implications for researchers, practitioners, and educators, and by highlighting priorities for stronger evidence—particularly external validation, transparent leakage safeguards, and evaluation aligned with clinical purpose.

KEYWORDS: Systematic Literature Review, Random Forest, Under-five children, Risk prediction, Model validation.

1. INTRODUCTION

1.1 *background of the study*

Machine learning is revolutionising several sectors, including child health and development, by providing innovative solutions to challenging issues. Specifically, Random Forest supervised machine learning algorithm which is a “tree-based” technique in which several decision trees are constructed from a random set of covariates and used to predict an outcome label for a subset of samples (Fenta et al., 2021). The algorithm is the ensemble of a large number of individual decision trees, and the final classification or prediction is obtained from the class with the most votes (Okagbue et al., 2021). It functions by creating and combining several decision trees derived from various bootstrap samples of the dataset, which helps to enhance predictive accuracy and reduce the risk of overfitting (Mthethwa & Melesse, 2025). Additionally, one of the major key features of the random forest is that it is able to deal with high-dimensional datasets and handle missing data with minimal preprocessing. Another notable strength of Random Forest is its capability to estimate feature importance (Mthethwa & Melesse, 2025).

Within under-five health and development monitoring, Random Forest has been repeatedly adopted because it performs reliably on the dominant data structures in this area—particularly mixed tabular predictors combining clinical, anthropometric, household, and sociodemographic factors—while also offering a practical pathway for identifying influential determinants through variable-importance outputs (Fenta et al., 2021). Alongside survey-style prediction, recent work extends Random Forest to engineered feature representations derived from chest X-ray images and respiratory audio. In pneumonia detection from chest X-rays, Random Forest is commonly

embedded within classical machine-learning pipelines built on handcrafted or engineered features rather than end-to-end deep learning, sometimes reporting very high internal performance (Panda & Das, 2025; Brindha et al., 2024). In respiratory sound screening, however, Random Forest can exhibit a clinically important precision–recall trade-off when applied to high-dimensional embeddings, where high precision may coincide with low sensitivity for the positive class (Ehtesham et al., 2025). This variation highlights that Random Forest performance cannot be interpreted meaningfully without attention to the clinical purpose (screening versus diagnosis), the representation type (tabular versus embeddings), and the preferred error profile. Evidence from population-level risk prediction further suggests that Random Forest is frequently competitive but not consistently optimal when compared with well-tuned boosting-based ensembles. In Acute respiratory infections (ARI) prediction using Bangladesh Demographic and Health Survey (BDHS) microdata, Random Forest performs strongly after imbalance handling but can be surpassed by gradient boosting methods (Hasan et al., 2024). By contrast, in NMIS-based malaria prediction, Random Forest is reported as a comparatively balanced model when evaluated across discrimination metrics and confusion-matrix trade-offs (Mthethwa & Melesse, 2025). In nutrition and growth monitoring, Random Forest appears particularly persuasive when outcomes align with established measurement rules and clinically proximal predictors, including anthropometric indices and indicators of severity (Fenta et al., 2021; Momand et al., 2022). Across developmental monitoring tasks, Random Forest is widely used as a baseline and, increasingly, is situated alongside explainable modelling approaches that aim to make

risk factors transparent for policy and practice (Hossain et al., 2023; Mim et al., 2025).

Despite this breadth of use, the literature remains fragmented. Studies differ substantially in dataset scale (large national surveys versus small hospital cohorts), data modality (tabular versus image/audio representations), imbalance handling (e.g., SMOTE or oversampling), validation design (simple hold-out splits versus cross-validation, and how tuning is separated from evaluation), and metric choice (accuracy versus AUROC and recall-focused reporting). These differences restrict cross-study comparability and make it difficult to generalise claims about the conditions under which Random Forest should be preferred, the reliability of reported performance, and the extent to which variable-importance results are stable and transferable across contexts.

Accordingly, this systematic literature review synthesises the evidence on Random Forest for monitoring health and development outcomes in children under five years. It aims to contribute to the body of knowledge already available on the application of Random Forest machine learning algorithm in children under-five:

- RQ1- How is Random Forest operationalised in under-five health and development monitoring systems?
- RQ2- What data regimes and input representations are used in Random Forest-based under-five monitoring?

1.2 Objectives

This systematic literature review aims to consolidate and critically synthesise evidence on the use of Random Forest for monitoring health and development outcomes in children under five years. Specifically, the review seeks to map how Random Forest has been deployed across diverse endpoint families—respiratory conditions, malaria, nutrition

and growth outcomes, and developmental monitoring—and to categorise its recurring roles as a primary predictive model, a benchmark comparator within multi-model evaluations, and a determinant-oriented tool that supports interpretation through variable-importance ranking. In doing so, the review also characterises the dominant data regimes underpinning these applications, including large-scale national survey microdata, smaller clinical and hospital datasets, and high-dimensional sensor or repository modalities such as chest X-ray images and respiratory audio.

A second objective is to evaluate the methodological quality and comparability of the Random Forest evidence base by examining dataset size, validation design, and evaluation metrics, with particular attention to how class imbalance is handled. The review therefore analyses the extent to which reported model performance is shaped by differences in hold-out splits versus cross-validation, separation of tuning from evaluation, and the use of resampling techniques such as SMOTE or oversampling. By linking these methodological choices to the observed performance patterns—especially clinically consequential error profiles in screening contexts (e.g., sensitivity/recall)—the review aims to identify robust conclusions, clarify where evidence may be optimistic or context-bound, and propose future research priorities such as external validation, distribution-shift testing, calibration and threshold reporting, and stability analysis of variable-importance results.

2. MATERIALS AND METHODS

2.1 Research Design

To achieve the objectives mentioned above, drawing on available suggested steps and recommendations for reporting (Moher et al., 2009), this work adopts the PRISMA in conducting a

systematic literature review (SLR) of Random Forest machine learning algorithm application for children under-five. The research took place on 28/2/2026.

First, this study identified the related articles "focuses mainly on peer-reviewed research articles published in reputable journals or conference proceedings, which specifically address Random Forest application for children under-five. Scopus databases had been used through relevant search terms, and within specific scope of computer science, engineering, & social science. Prior comparative studies have demonstrated that Scopus indexes a larger number of peer-reviewed journals and provides more comprehensive citation coverage, especially for journals published outside North America and Western Europe (Falagas et al., 2008; Mongeon & Paul-Hus, 2016). Mongeon and Paul-Hus (2016), in a large-scale analysis published in Scientometrics, empirically showed that Scopus covers approximately 60–70% more journals than WoS, with particularly strong representation in engineering, computer science, and business-related disciplines. Given the inherently multidisciplinary nature of Machine Learning algorithm (i.e., Random Forest), the use of Scopus is therefore methodologically justified and consistent with systematic studies in this research area. Next, the article titles and abstracts were screened to identify potentially eligible studies according to the inclusion criteria. Finally, eligible articles were investigated thoroughly for further synthesis.

2.2 Inclusion criteria

The inclusion criteria were articles that are (a) written in both English, and Bahasa Indonesia, (b) published anytime in the scope of computer science, engineering, and social

science, and finally, (c) it matches the search string:

(TITLE-ABS-KEY ("Random Forest") AND TITLE-ABS-KEY ("child") AND TITLE-ABS-KEY (classification) AND TITLE-ABS-KEY (under-five)) AND (LIMIT-TO (SUBJAREA , "COMP") OR LIMIT-TO (SUBJAREA , "ENGI") OR LIMIT-TO (SUBJAREA , "SOCI")). The studies included in this review focused on the application of Random Forest algorithm for children under-five scope. The inclusion criteria were: studies that employed Random Forest algorithm, studies that reported quantitative performance metrics, and studies published in peer-reviewed journals. Studies were excluded if they did not involve Random Forest algorithm in the developed model in the study, or were not published in peer-reviewed journals.

2.3 Screening Procedure and Eligibility Assessment

The retrieved articles were brought using Scopus databases were 18 articles. A discussion within the research team has done further investigation for the retrieved articles. Firstly, a thorough review of the articles title, abstract, and keywords conducted based on the inclusion of Random Forest, children, under-five keywords, & there were 18 included articles as the last outcome for further investigation. Finally, one retrieved paper where a book chapter and it was inaccessible, two papers were non-related as one of it did not implement Random Forest, while the second were a pure data mining not machine learning, thus remaining 15 papers for further investigation. (see Figure 1 for a PRISMA flowchart of the entire procedure).

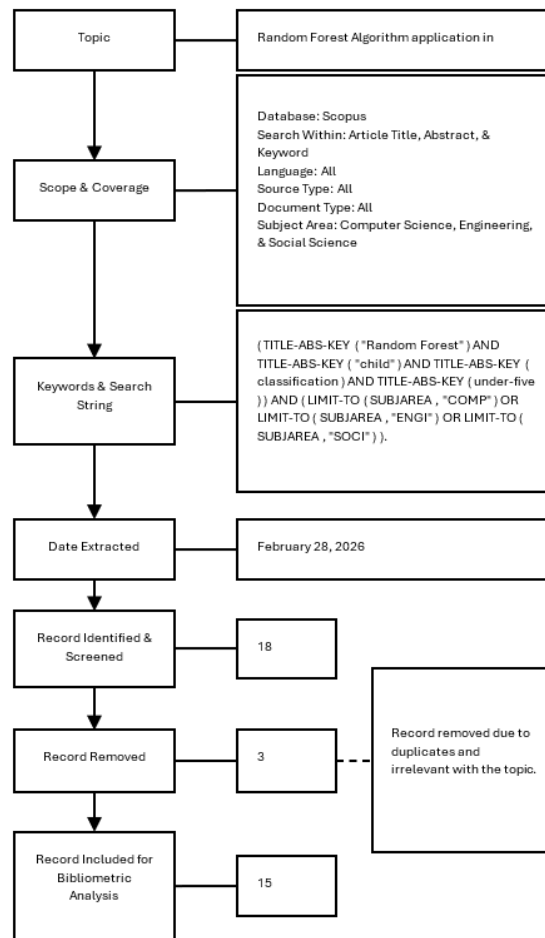


Figure 1 Flowchart of systematic literature review, adapted from Moher et al. (2009).

3. RESULTS AND FINDINGS

Across the 15 included studies, Random Forest is consistently presented as a practical “workhorse” classifier rather than a purely theoretical option. It is repeatedly chosen because it performs robustly on the data types that dominate child-health and child-development research—especially mixed tabular inputs that combine clinical, anthropometric, and sociodemographic variables. More recently, it is also being applied to engineered feature representations extracted from signals or images, where it remains a dependable baseline and, in some cases, a strong primary model.

3.1 Data Extraction and Synthesis

Across the included studies, Random Forest was used in three recurring roles within under-five monitoring tasks: (i) as the primary classifier, (ii) as a

benchmark comparator in multi-model evaluations, and (iii) as a feature-importance or variable-ranking tool to support determinant-focused interpretation. The researchers organised the evidence by endpoint family—respiratory infection, malaria, nutrition/growth, and developmental monitoring—and summarised each group in terms of data modality (tabular survey/clinical records versus engineered features from images or audio), dataset scale, validation approach, and the evaluation metrics reported.

3.2 Type of System

Across the respiratory studies, RANDOM FOREST is mainly used as a classical ML classifier, but its behaviour varies sharply depending on the input representation and the screening objective.

In CXR-based pneumonia work, RANDOM FOREST is embedded in engineered-feature pipelines rather than end-to-end deep learning. One edge-feature approach reports 98.72% accuracy after explicit class balancing (oversampling from 5,863 to 8,340 images) and feature fusion (e.g., Canny–Sobel). A second CXR workflow—built around simpler image vectorisation for non-CNN classifiers—reports RANDOM FOREST outperforming baseline models such as decision tree and kNN, with 92.9% accuracy. In audio-based pneumonia screening, respiratory sounds are converted into 512-dimensional embeddings derived from short (2-second) clips and then classified using standard ML models including RANDOM FOREST. Here, RANDOM FOREST shows a clinically problematic error profile: it achieves high positive-class precision (0.90) but very low positive-class recall (0.13), with the confusion matrix dominated by false negatives—an outcome that is difficult to accept in a screening context where sensitivity is typically prioritised.

For ARI prediction using survey microdata, RANDOM FOREST is evaluated alongside multiple classifiers on a BDHS sample of 11,157 under-five children with 2.5% prevalence. After SMOTE-based balancing and missingness handling, RANDOM FOREST reaches 0.889 accuracy, but XGBoost performs better at 0.910, reinforcing the broader pattern that boosting-based ensembles can surpass RANDOM FOREST in large tabular prediction tasks.

In NMIS 2021 malaria prediction ($n = 10,655$ under-five children tested by RDT), RANDOM FOREST is presented as the most balanced model. It achieves 0.79 accuracy and 0.80 AUC, with the confusion-matrix results described as showing a favourable trade-off between false positives and false negatives compared with competing approaches. Finally, in the nutrition and growth domain, RANDOM FOREST appears most persuasive when

predictors are clinically proximal and the outcome aligns closely with established measurement rules. In undernutrition determinant modelling using EDHS-scale data ($n = 29,333$ children aged 0–59 months), RANDOM FOREST is reported as the strongest model ($\text{AUC} \approx 0.761$) and is also used to rank covariates through its importance measures. In hospital-based oedematous malnutrition severity prediction, RANDOM FOREST performance increases substantially after SMOTE, with reported post-SMOTE accuracy reaching 97.14%, illustrating both RANDOM FOREST's ability to exploit strong clinical signal and the sensitivity of reported performance to resampling design in imbalanced clinical datasets.

3.3 Data sources

The evidence base falls into three main data regimes. First, nationally representative survey microdata dominate population-level risk prediction, typically with large samples (roughly 10,000–30,000 observations) and recurring applications to malaria, ARI, undernutrition determinants, and mortality-related tasks. Second, clinical and hospital datasets are usually smaller and site-specific, and are more often used for severity classification and decision-support settings where predictors are closely tied to clinical workflows. Third, repository- or sensor-like modalities—notably CXR images and respiratory audio—support high-dimensional inputs, either as engineered feature sets or foundation-model embeddings, paired with clinically annotated labels.

3.4 Validation Designs and Evaluation Metrics

Most studies relied on internal validation, typically using a hold-out split (often 80/20) and, less consistently, cross-validation. Where cross-validation was used, its purpose varied: some papers applied CV mainly for hyperparameter

tuning, while others incorporated repeated CV into the main evaluation and then reported performance again on a separate held-out test set.

Evaluation metrics also varied with the endpoint and the degree of class imbalance. Several studies reported AUROC, sensitivity/specificity, precision/recall, F1, and agreement statistics such as Cohen's kappa, whereas others emphasised accuracy as the headline result. In strongly imbalanced problems (for example, ARI prevalence of 2.5%), many studies explicitly applied resampling strategies—most commonly SMOTE or random oversampling—to rebalance the training data prior to model fitting.

4. DISCUSSION and FUTURE WORKS

Across endpoints, a clear pattern is that RANDOM FOREST is repeatedly chosen because it performs robustly on heterogeneous, mixed-type predictors and can support both prediction and determinant-style interpretation through its internal importance measures. However, the comparative evidence suggests that RANDOM FOREST is more often competitive than consistently optimal—especially in large, tabular benchmark settings where well-tuned boosting methods are included. Taken together, this supports positioning RANDOM FOREST as a strong, interpretable benchmark model, rather than treating it as the default best-performing choice.

The most consequential source of heterogeneity across studies is data modality. In CXR-based pipelines, RANDOM FOREST often benefits from engineered feature sets that align well with tree-based partitioning, and this can translate into very strong internal performance. By contrast, in embedding-driven audio pipelines, RANDOM FOREST may default to a more conservative

decision boundary—seen in high precision but low recall—unless it is explicitly optimised for sensitivity and imbalance-related costs. This reinforces a key implication for under-five screening: performance should not be judged on accuracy alone. The clinically relevant question is often whether the model meaningfully reduces missed cases, given the high cost of false negatives in screening-oriented deployment.

Cross-study comparability is limited by three recurring methodological differences. First, studies vary in validation rigour—including how train-test splits are constructed, whether repeated cross-validation is used, and whether hyperparameter tuning is clearly separated from final evaluation. Second, comparability is affected by metric choice: papers that foreground accuracy are not directly comparable to those that emphasise AUROC or recall/sensitivity, particularly for screening endpoints. Third, results are sensitive to imbalance handling, as SMOTE or oversampling can substantially alter the reported performance profile.

As a result, exceptionally large post-resampling gains should be interpreted cautiously unless studies transparently report leakage controls—such as patient-level separation where relevant, resampling performed only within training folds, and nested cross-validation (or equivalent procedures) to prevent optimistic bias.

Future work can strengthen the evidence base by designing models around the clinical purpose rather than treating performance as a single, generic target. For screening applications, the evaluation should prioritise sensitivity/recall, supported by threshold calibration and transparent reporting of the resulting false-negative burden. For determinant-oriented monitoring, studies should go beyond point estimates of importance and report the

stability and robustness of variable rankings (for example, across resamples, folds, or alternative specifications).

More broadly, stronger claims about real-world utility will depend on external validation and explicit distribution-shift testing. This is particularly important for repository-based imaging pipelines and device-dependent audio systems, where differences in acquisition conditions, hardware, and clinical settings can materially change model behaviour even when internal validation appears strong.

5. CONCLUSION

This systematic literature review targeted Scopus-indexed, peer-reviewed studies that applied Random Forest methods to under-five child health across computing, engineering, and social science disciplines, retrieving English and Bahasa Indonesia articles that met the predefined search string and inclusion criteria. Fifteen studies were eligible, deploying Random Forests for respiratory infection surveillance, malaria risk prediction, nutrition and growth monitoring, and developmental assessment using national survey datasets, clinical records, and imaging/audio repositories. Reported model performance was frequently encouraging and several works demonstrated prototype web deployments, yet the corpus exhibits substantive methodological limitations: evaluation metrics were often mismatched to clinical intent, external validation was infrequent, potential data-leakage pathways were insufficiently mitigated, and stability of variable-importance rankings was rarely assessed. For Random Forest applications to mature into reliable, generalisable tools for under-five health, future research must align model design and metric selection with explicit clinical objectives (e.g.,

prioritise sensitivity and calibrate thresholds for screening), routinely quantify the robustness of determinant rankings across resamples and model specifications, and undertake external validation and distribution-shift testing—especially for models built on heterogeneous imaging and device-dependent audio data.

ACKNOWLEDGMENT(S)

The author(s) would like to thanks to all members of this work including the research team, the developers and the university staff. This research was funded by Thailand Science Research and Innovation (TSRI).

REFERENCES

- Brindha, M., Maheswari, S., Vidhyabharathi, S., & Yaspe Francis, M. V. (2024). Utilizing Random Forest to Diagnose Pneumonia in Children. 2024 3rd International Conference on Artificial Intelligence for Internet of Things, Allot 2024, (Allot), 1–6. <https://doi.org/10.1109/Allot58432.2024.10574685>
- Ehtesham, A., Singh, A., Kumar, S., & Khoei, T. T. (2025). Early Detection of Pediatric Pneumonia Using Google's HeAR Model: A Respiratory Sound Embedding Approach. 2025 IEEE 6th Annual World AI IoT Congress, Allot 2025, 185–191. <https://doi.org/10.1109/Allot65859.2025.11105317>
- Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., & Pappas, G. (2008). Comparison of PubMed, Scopus, Web of Science, and Google Scholar: Strengths and weaknesses. *FASEB Journal*, 22(2), 338–342. <https://doi.org/10.1096/fj.07-9492LSF>
- Fenta, H. M., Zewotir, T., & Muluneh, E. K. (2021). A machine learning classifier approach for identifying the determinants of under-five child undernutrition in Ethiopian administrative zones. *BMC Medical Informatics and Decision Making*, 21(1), 1–12. <https://doi.org/10.1186/s12911-021-01652-1>
- Hasan, M. N., Emon Proddhan, M., Alam, A., Sultana, S., Rahman, M., & Parvin, H. (2024). A Comparative Study of Machine Learning Algorithms in Predicting Acute Respiratory Infection of Under-Five Children in Bangladesh Using Imbalanced Data. 2024 IEEE International Conference on Biomedical Engineering, Computer and Information Technology for Health, BECITHCON 2024, (November), 274–278. <https://doi.org/10.1109/BECITHCON64160.2024.10962602>
- Hossain, M. I., Haq, I., Talukder, A., Suraiya, S., Rahman, M., Saleheen, A. A. S., Methun, M. I. H., Habib, M. J., Hossain, M. S., Nayan, M. I. H., & Hussain, S. (2023). Performance Evaluation of Machine Learning-Based Algorithms to Predict the Early Childhood Development Among Under Five Children in Bangladesh.

- Journal of Computer Science, 19(5), 641–653.
<https://doi.org/10.3844/jcssp.2023.641.653>
- Husaini, A., Hoeronis, I., Lumana, H. H., & Pusporeni, L. D. (2023). Early Detection of Stunting in Toddlers Based on Ensemble Machine Learning in Purbaratu Tasikmalaya. *Jurnal Sistem Dan Teknologi Informasi (JustIN)*, 11(3), 487. <https://doi.org/10.26418/justin.v11i3.66465>
- Islam, M., Usman, M., Mahmood, A., Abbasi, A. A., & Song, O. Y. (2020). Predictive analytics framework for accurate estimation of child mortality rates for Internet of Things enabled smart healthcare systems. *International Journal of Distributed Sensor Networks*, 16(5). <https://doi.org/10.1177/1550147720928897>
- Maasthi, M. J., Gururaj, H. L., Ravi, V., D. B., Almeshari, M., & Alzamil, Y. (2023). Decision-making Support System for Predicting and Eliminating Malnutrition and Anemia. *The Open Bioinformatics Journal*, 16(1), 1–11. <https://doi.org/10.2174/0118750362246898230921054021>
- Mim, M. A., Khatun, M. R., Hossain, M. M., Rahman, W., & Munir, A. (2025). Exploring Early Learning Challenges in Children Utilizing Statistical and Explainable Machine Learning. *Algorithms*, 18(1). <https://doi.org/10.3390/a18010020>
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group*. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of internal medicine*, 151(4), 264-269 should be updated to "Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Bmj*, 339.
- Momand, Z., Zarinkhail, M. S., & Aryan, M. F. (2022). Machine Learning Based Prediction of Edematous Malnutrition in Afghan Children. *Lecture Notes in Networks and Systems*, 322, 235–245. https://doi.org/10.1007/978-3-030-85990-9_20
- Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: A comparative analysis. *Scientometrics*, 106(1), 213–228. <https://doi.org/10.1007/s11192-015-1765-5>
- 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), 1–12. <https://doi.org/10.2174/0118749445396163250604103305>
- 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>
- Panda, P., & Das, S. (2025). Paediatric Pneumonia Detection Using Edge-Based Feature Extraction Techniques. *International Conference on Artificial Intelligence and Emerging Technologies, ICAIET 2025*, 1–6. <https://doi.org/10.1109/ICAET65052.2025.11211131>
- Ziyad, S. R., Liyakathunisa, Aljohani, E., & Saeed, I. A. (2023). Diagnosis of Autism Spectrum Disorder by Imperialistic Competitive Algorithm and Logistic Regression Classifier. *Computers, Materials and Continua*, 77, 1515–1534. <https://doi.org/10.32604/cmc.2023.040874>