Research Square

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

Classifying Dengue with Warning Signs Using Routine Clinical Features: A Machine Learning Study in a Nepalese Cohort

Chandramani Wagle
Global Clinical Research
Dinesh Prasad Ghimire
Global Clinical Research
Abhay Kumar Sah
Global Clinical Research
Nikita Acharya
Global Clinical Research
Mahesh Lamsal
Virology Society Nepal
Apshara Parajuli
Virology Society Nepal
Sandeepan Bhatia
Global Clinical Research
Surendra Urnaw
B.P. Koirala Institute of Health Science
Punam Yadav
Cistron Diagnostic Reliable Lab
Huta Banjade
Independent Researcher
Birendra Prasad Gupta
birendraphd@gmail.com

Virology Society Nepal

Research Article

Keywords: Dengue, Machine learning, Severity prediction, XGBoost, Random Forest, Nepal

Posted Date: May 21st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9452907/v1

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

Additional Declarations: No competing interests reported.

Abstract

Background

Dengue is a major public health concern in tropical and subtropical regions, with increasing incidence and expanding geographic distribution. Early identification of patients requiring closer monitoring and clinical attention remains a clinical challenge, particularly in resource-limited settings where access to advanced diagnostic tools is constrained. This study aimed to develop and evaluate machine learning models for classifying dengue with warning signs (DWS) versus dengue without warning signs (DWWS).

Methods

A hospital-based, multi-site retrospective study was conducted using data from 1,618 dengue patients, of whom 1,601 were included after data preprocessing. Predictor variables included demographic factors, comorbidities, and symptom severity scores derived from clinical assessments. Eight supervised machine learning algorithms (Logistic Regression, Random Forest, XGBoost, Support Vector Machine, Naïve Bayes, Decision Tree, Multi-Layer Perceptron, and AdaBoost) were trained and evaluated. Data were split into training (80%) and testing (20%) sets using stratified sampling. Model performance was assessed using area under the receiver operating characteristic curve (AUC-ROC), accuracy, and precision–recall metrics. The study was conducted in accordance with the Declaration of Helsinki and ethical approval was received from the Ethical Review Board (ERB) of the Nepal Health Research Council (NHRC), approval number 45.

Results

Ensemble-based models demonstrated superior performance compared to other algorithms. XGBoost achieved the highest discrimination with an AUC-ROC of 0.835, followed closely by Random Forest (AUC-ROC: 0.831), which also showed the highest accuracy (0.78). Other models, including Support Vector Machine and Decision Tree, showed moderate performance, while Logistic Regression and Naïve Bayes demonstrated comparatively lower predictive ability. Precision–recall analysis confirmed the robustness of ensemble models, with performance exceeding the baseline prevalence levels. Feature importance analysis identified chest pain, retro-orbital pain, muscle pain, and conjunctivitis as the most influential predictors.

Conclusions

Machine learning models, particularly ensemble-based approaches, show strong potential for classifying dengue cases based on warning sign status using routinely available clinical data at the time of presentation. These models may support early risk stratification and clinical decision-making, especially

in resource-limited settings. Further validation using external datasets and real-world implementation studies is needed before clinical integration.

Background

Dengue has emerged as one of the fastest-growing vector-borne viral diseases globally, with incidence rising dramatically over the past two decades [1]. Cases reported to the World Health Organization (WHO) increased from 505,430 in 2000 to an unprecedented 14.6 million in 2024, with more than 12,000 dengue-related deaths in the same year [1, 2]. The disease is now endemic in over 100 countries, placing an estimated 5.6 billion people at risk of infection [1, 3]. Dengue affects tropical and subtropical regions worldwide, including Southeast Asia (SEA), the Americas, Africa, and the Pacific Islands, and is increasingly reported in parts of Europe and the Eastern Mediterranean [1, 4]. South and Southeast Asia account for over half of the global dengue burden; in 2020, of the 3.5 billion people living in endemic countries, 1.3 billion resided in just 10 SEA nations, with India, Indonesia, Myanmar, Sri Lanka, and Thailand among the most affected [5, 6].

Nepal has faced increasingly severe dengue outbreaks recently, with the largest recorded in 2022 affecting all seven provinces and 77 districts. Outbreaks have also become more frequent and intense, with consecutive epidemics occurring between 2022 and 2024. Subsequent epidemics in 2023 and 2024 caused over 85,000 reported infections and more than 33 verified deaths [7, 8]. These outbreaks are not confined to traditional dengue-prone areas; they are expanding geographically, affecting regions previously considered dengue-free [9]. This geographic expansion places additional strain on Nepal's already limited healthcare infrastructure, which faces critical shortages of hospital beds and medical personnel, particularly in rural areas [10, 11].

The WHO 2009 dengue classification, categorizing cases as dengue without warning signs (DWWS), dengue with warning signs (DWS), and severe dengue (SD), serves as the global standard for clinical management and triage [12]. While widely adopted, this framework faces practical challenges, as dengue often presents with overlapping symptoms and variable clinical manifestations that complicate early risk stratification [13, 14]. Accurate categorization is further constrained by reliance on clinician judgment and the absence of standardized tools, which may lead to both underestimation and overestimation of disease severity. Moreover, conventional clinical approaches may overlook subtle interactions between clinical features and laboratory findings that could better predict disease progression [14, 15]. These limitations emphasize the necessity of data-driven approaches capable of integrating diverse patient information to improve early detection of severe cases and support evidence-based decision-making.

Machine learning (ML) has gained increasing attention in infectious disease research for its capacity to model complex, nonlinear interactions among diverse clinical and laboratory variables [16, 17]. In dengue, ML applications from countries such as India, Thailand, and Brazil have demonstrated promising results in predicting disease severity and clinical outcomes [18]. However, most existing models rely

heavily on advanced laboratory parameters or imaging data, which may not be readily available in resource-limited settings. Importantly, despite Nepal's recurrent nationwide dengue outbreaks, to our knowledge, there is limited published evidence on the application of ML approaches to predict dengue severity in Nepalese patient cohorts. This gap underscores the need for context-specific, data-driven models that leverage routinely available clinical features for early risk stratification.

In this study, we applied and compared supervised machine learning algorithms to classify dengue severity as DWS versus DWWS, using routinely collected clinical and demographic features from patients in Nepal. To our knowledge, this is among the first studies in Nepal to apply ML-based approaches for dengue severity prediction, distinguished by its focus on routinely available clinical features rather than reliance on advanced laboratory diagnostics or imaging. By evaluating a diverse set of models and examining feature contributions, we aimed to lay the groundwork for interpretable, clinically useful ML tools that can support early risk stratification of dengue patients.

Methods

Study Design and Settings

This hospital-based, multi-site study involved a retrospective analysis of prospectively collected dengue patient data during the outbreak years of 2016, 2017, 2019, and annually from 2021 through 2024.

In 2016–2017, data were obtained from Chitwan Medical College (Chitwan) and Bhawani Hospital (Birgunj). For 2019 and 2021–2024, data were obtained from Bijayapur Hospital, Sunkoshi Laboratory, and B.P. Koirala Institute of Health Sciences (BPKIHS), all located in Dharan, and from Itahari Municipal Hospital (Itahari). These facilities are key referral centers serving diverse urban and semi-urban populations in eastern Nepal. Ethical approval was obtained from the Nepal Health Research Council in Kathmandu. (Approval number 45)

Study Population and Eligibility

The study population comprised patients diagnosed with dengue infection who presented to the outpatient departments (OPDs) of the participating hospitals during the study period. Eligibility criteria included individuals with a fever of ≤ 72 hours and clinical symptoms suggestive of dengue, who were confirmed by positive NS1 antigen or IgM/IgG serology. Exclusion criteria included records with missing data that precluded classification under the 2009 WHO dengue guidelines, patients with clinical features suggestive of alternative diagnoses (e.g., pneumonia, otitis media), and individuals with chronic conditions such as malignancy, advanced liver or kidney disease, or immunosuppression. Pregnant women were excluded to minimize confounding, as pregnancy is an independent risk factor for severe dengue outcomes.

Data Collection

Data were collected at each study site using a structured case report form (CRF) developed by the research team in consultation with clinical and epidemiological experts (Supplementary File 1). Participation was voluntary, and written informed consent was obtained from participants at the time of data collection. For participants < 18 years of age, written informed consent was obtained from a parent or legal guardian. Trained research assistants conducted face-to-face interviews and reviewed medical records to collect demographic, clinical, and laboratory information. Completed CRFs were reviewed daily by site supervisors for completeness and consistency. No personally identifiable information was collected to ensure participants' confidentiality.

Dengue Severity Classification

Dengue severity was classified in two sequential steps prior to model development to ensure clear outcome labels and avoid label leakage.

First, records were screened for SD based on available clinical information. Cases were labeled as SD and excluded from modeling if any of the following were present: shock, respiratory distress with fluid accumulation, severe bleeding, alanine aminotransferase ≥ 1000 U/L, central nervous system involvement, or organ failure.

Second, the remaining cases were classified as DWS or DWWS according to the 2009 WHO criteria, using binary clinical indicators recorded at enrollment [12]. This classification aligns with Nepal's guidelines for managing dengue cases [19]. DWS was defined by the presence of at least one warning sign, including abdominal pain or tenderness, persistent vomiting, clinical fluid accumulation, mucosal bleeding, or lethargy/restlessness. Hepatomegaly > 2 cm was not consistently available and was not included. Cases without these features were classified as DWWS. As only one patient met the SD criteria, this case was removed. Thus, the final modeling outcome was defined as a binary classification of DWS versus DWWS.

Data Preprocessing and Feature Engineering

All analyses were conducted using Python 3 within the Anaconda environment, with Jupyter Notebook as the interface. Machine learning pipelines were implemented using the scikit-learn library and related scientific computing packages.

Data preprocessing included encoding, handling missing data, class balancing, and normalization. Categorical and ordinal variables were converted into numerical labels. Missing data were minimal (approximately 1%; 16 of 1,617 records) and were removed rather than imputed, resulting in a final dataset of 1,601 records with 16 variables.

To address class imbalance, random oversampling of the minority class (DWS) was performed within the training set while preserving the test set distribution. Continuous variables were standardized (zero mean and unit variance) for models sensitive to feature scaling, including Logistic Regression (LR), Support Vector Machine (SVM), and Multi-Layer Perceptron (MLP). Standardization was also applied for consistency to Gaussian Naïve Bayes (GNB), although it is not strictly required. No scaling was applied to ensemble methods (Random Forest, XGBoost, and AdaBoost), particularly tree-based models such as RF and XGBoost.

Fifteen predictor variables were selected based on clinical relevance and included demographic factors (sex), pre-existing conditions (diabetes and obesity, defined as body mass index (BMI) ≥ 30 kg/m²), and symptom severity scores derived from clinical features (fever, headache, nausea, retro-orbital pain, joint pain, malaise, muscle pain, rash, diarrhea, flushing, chest pain, and conjunctivitis). These scores were used to capture the intensity and presence of symptoms in a structured and quantifiable manner. Symptom severity scores were derived from structured CRF data based on the presence and reported intensity of each symptom using predefined scoring criteria applied consistently across study sites. The final dataset showed a moderate class imbalance, with DWS cases comprising approximately 43.4% of the total sample.

Model Development and Hyperparameter Tuning

Eight supervised machine learning algorithms were trained and evaluated: LR, RF, XGBoost, SVM, GNB, Decision Tree (DT), MLP, and AdaBoost. These models were selected due to their established use in clinical datasets and their ability to capture non-linear relationships while handling modest sample sizes.

The dataset was split into training (80%) and testing (20%) sets using stratified sampling to preserve class distribution. Hyperparameter tuning was performed using grid search with 20 iterations of stratified shuffle-split cross-validation on the training set. Random seeds were fixed to ensure reproducibility.

Model Evaluation and Performance Metrics

Model performance was evaluated on the held-out test set using multiple metrics, including accuracy, precision, recall, F1-score, area under the receiver operating characteristic curve (AUC-ROC), and area under the precision–recall curve (AUC-PR).

Performance metrics were derived from the confusion matrix, where predictions were categorized as true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). In this study, a positive prediction corresponds to DWS, while a negative prediction corresponds to DWWS.

The evaluation metrics were calculated as follows:

- Accuracy = (TP + TN) / (TP + TN + FP + FN)

- Precision = TP / (TP + FP)
- Recall = TP / (TP + FN)
- F1-score = 2 × (Precision × Recall) / (Precision + Recall)

AUC-ROC and AUC-PR were used as threshold-independent measures of model discrimination, with AUC-PR particularly informative under class imbalance. Final model selection was based on overall discrimination performance, primarily AUC-ROC and AUC-PR.

Results

Study Population and Dataset Characteristics

A total of 1,617 patient records were initially available. After data preprocessing, 16 records containing invalid or non-numeric entries were excluded, resulting in a final analytic dataset of 1,601 patients. The demographic and clinical characteristics of the study population are summarized in Table 1. The cohort comprised 785 males (49.0%). Pre-existing diabetes was reported in 108 patients (6.7%), while 53 patients (3.3%) were classified as obese. Among the final dataset, DWS cases accounted for 695 (43.4%) patients, while the remaining were classified as DWWS.

Table 1. Baseline Demographic and Clinical Characteristics of the Study Population (n = 1,601)

Variable	Overall
Demographic characteristics, n (%)	
male	785 (49%)
Female	816 (51%)
Comorbidities, n (%)	
Diabetes	108 (6.7%)
Obesity (BMI ≥ 30 kg/m²)	53 (3.3%)
Clinical features, mean ± SD	
Fever score	5.29 ± 4.12
Headache score	10.89 ± 10.80
Nausea score	6.81 ± 7.53
Retro-orbital pain score	8.65 ± 9.83
Joint pain score	9.79 ± 11.79
Malaise score	7.68 ± 9.45
Muscle pain score	9.80 ± 12.85
Rash score	1.59 ± 4.20
Diarrhea score	1.38 ± 7.37
Flushing score	1.00 ± 3.99
Chest pain score	0.42 ± 2.46
Conjunctivitis score	0.70 ± 3.93

Clinical features were represented as symptom severity scores. Among these, headache (mean ± SD: 10.89 ± 10.80), joint pain (9.79 ± 11.79), and muscle pain (9.80 ± 12.85) had the highest average scores, indicating greater symptom burden in these domains. Moderate mean scores were observed for retro-orbital pain (8.65 ± 9.83), malaise (7.68 ± 9.45), and nausea (6.81 ± 7.53). Lower mean scores were noted for rash (1.59 ± 4.20), diarrhea (1.38 ± 7.37), and flushing (1.00 ± 3.99), while chest pain (0.42 ± 2.46) and conjunctivitis (0.70 ± 3.93) were the least prominent clinical features in the study population. Fever score demonstrated a moderate distribution (5.29 ± 4.12).

Model Performance Comparison

Eight supervised machine learning models were trained using the selected clinical and demographic predictors, and their performance was evaluated on a held-out test set. The comparative performance of all models is presented in Table 2.

Table 2. Model performance with Area Under the Curve (AUC-ROC) and Accuracy

Model	AUC-ROC	Accuracy
Random Forest (RF)	0.831	0.78
Extreme Gradient Boosting (XGBoost)	0.835	0.77
Decision Tree (DT)	0.759	0.72
AdaBoost (ADB)	0.728	0.71
Logistic Regression (LR)	0.715	0.70
Support Vector Machine (SVM)	0.779	0.70
Neural Network (Multi-Layer Perceptron, MLP)	0.70	0.68
Naïve Bayes (NB)	0.685	0.57

AUC-ROC represents the model's ability to discriminate between outcome classes across all classification thresholds. Accuracy reflects the proportion of correctly classified cases in the test set.

Among the evaluated models, ensemble tree-based methods demonstrated superior predictive performance. The XGBoost model achieved the highest discrimination ability, with an AUC-ROC of 0.835 and an accuracy of 0.77. The RF model showed comparable performance, with an AUC-ROC of 0.831 and the highest accuracy (0.78) among all models. Other models demonstrated moderate performance. SVM achieved an AUC-ROC of 0.779, followed by the DT (0.759) and AdaBoost (0.728). LR showed fair discrimination (AUC-ROC: 0.715), while the Neural Network and NB models yielded lower predictive performance, with AUC-ROC values of 0.70 and 0.685, respectively (Table 2).

Overall, ensemble-based approaches consistently outperformed other evaluated models in this study across both discrimination and accuracy metrics, indicating their stronger capability to capture nonlinear relationships and higher-order feature interactions among clinical features. Figure 2 presents additional performance metrics, including precision, recall, and F1-score.

ROC and Precision–Recall Analysis

The receiver operating characteristic (ROC) curves for the evaluated models are presented in Figure 1. Consistent with the results in Table 2, ensemble-based models demonstrated superior discrimination performance. In particular, XGBoost and Random Forest showed the highest ability to distinguish between outcome classes across a range of decision thresholds.

Threshold-Based Performance

Beyond discrimination metrics, model performance was further evaluated using threshold-based measures on the held-out test set (Figure 2). Precision, recall, and F1-score were computed at a fixed decision threshold of 0.50 for probability-based models and 0.00 for models returning decision scores.

In addition to ROC-AUC, model discrimination was assessed using the area under the precision–recall curve (PR-AUC), which summarizes the trade-off between precision and recall across all possible thresholds. The PR-AUC baseline corresponds to the test-set prevalence (0.44), and values exceeding this baseline indicate improved predictive performance over a prevalence-level classifier.

As shown in Figure 2, ensemble-based models, particularly XGBoost and Random Forest, achieved the strongest performance across both PR-AUC and threshold-based metrics. These models demonstrated improved precision-recall trade-offs while maintaining balanced sensitivity and specificity at the default decision threshold. Differences between ROC-AUC and PR-AUC were observed across models, reflecting the influence of class distribution and the greater penalization of false positives in PR-based evaluation.

Overall, the consistency of model ranking across threshold-independent and threshold-dependent metrics supports the robustness of ensemble methods for classifying dengue warning signs.

Feature Importance

The contribution of individual predictors to model performance was assessed using feature importance derived from the XGBoost model (Figure 3). Feature importance was quantified using gain, which represents the average improvement in the model's loss function attributable to each feature during training. Gain values were normalized to represent the relative contribution of each feature to loss reduction during model training. As shown in Figure 3, chest pain emerged as the most influential predictor, followed by retro-orbital pain, muscle pain, and conjunctivitis. These features contributed substantially to the model's ability to distinguish between severity classes. In contrast, sex demonstrated the lowest relative importance, indicating a limited contribution to prediction in this context. The overall distribution of feature importance highlights the dominant role of specific clinical symptoms in driving model predictions, supporting the relevance of symptom-based features for dengue severity classification.

Discussion

In this multi-site study of 1,601 dengue patients in Nepal, ensemble tree-based methods (XGBoost and Random Forest) demonstrated superior discrimination for classifying the status of dengue warning signs (DWS vs. DWWS), achieving AUC-ROC values of 0.835 and 0.831, respectively, while conventional algorithms such as logistic regression and Naïve Bayes yielded substantially lower discrimination (AUC-ROC ≤ 0.715). These findings confirm that non-linear ensemble methods effectively capture complex interactions among routine clinical symptoms and demographic features, outperforming linear and

single-tree approaches on the held-out test set. These results align with and extend recent international evidence on machine learning for dengue severity prediction. Studies report that gradient-boosting and random-forest ensembles outperform logistic regression, support vector machines, and neural networks when applied to clinical or epidemiological datasets [20]. For instance, a recent study from Puerto Rico also found that gradient boosting algorithms outperformed other classifiers in predicting severe dengue [21].

The observed performance hierarchy, ensemble methods > SVM/decision > Logistic Regression/Naïve Bayes, reflects differences in algorithmic capacity to model nonlinear relationships and feature interactions in clinical data. Linear models such as LR and independence-based models such as Naïve Bayes demonstrated limited discrimination (AUC < 0.72), likely due to their limited ability to capture interaction effects between symptoms (e.g., chest pain and retro-orbital pain) without explicit feature engineering. This performance gap has important clinical implications: while simpler models offer interpretability advantages, their reduced accuracy may compromise patient safety in triage decisions. The moderate performance of neural networks (AUC 0.70) suggests that architectural complexity alone does not guarantee superior outcomes, particularly in modest sample sizes where data limitations and feature representation constrain performance, supporting the selection of tree-based ensembles as the optimal balance of accuracy and computational feasibility for this clinical context.

Feature importance analysis revealed chest pain as the dominant predictor (highest gain), followed by retro-orbital pain, muscle pain, and conjunctivitis, findings that are biologically coherent with established dengue pathophysiology. Chest pain in dengue often signals impending plasma leakage or myocardial involvement, both critical warning sign components under WHO 2009 guidelines [12]. Retro-orbital pain and myalgia are also classic dengue symptoms. If both of these symptoms are present, it could mean that the body is having a stronger inflammatory response and that blood vessels are more likely to leak. The minimal contribution of sex to prediction aligns with epidemiological evidence that, while dengue incidence may vary by gender in some settings, severity progression depends primarily on clinical factors rather than demographic characteristics [22]. These observations validate that the model's predictive logic aligns with clinical reasoning, enhancing trustworthiness for potential deployment. This supports the rationale for prioritizing symptom-based features when developing classification models to support early triage and risk stratification.

A key strength of this study is its practical design tailored for resource-limited settings. The proposed models rely exclusively on routinely available clinical symptoms, demographic characteristics, and comorbidities. They do not require laboratory investigations or imaging. This makes them well suited for outpatient and primary care use, particularly in dengue-endemic regions such as Nepal. The inclusion of multi-site data enhances the representativeness of the study population and improves the potential generalizability within similar healthcare contexts. In addition, the use of XGBoost with feature importance analysis provides a degree of interpretability, allowing identification of clinically relevant predictors and supporting clinician trust in model outputs. The growing recognition of interpretable machine learning approaches is crucial for promoting their adoption in clinical decision-making systems.

Despite these strengths, several limitations should be acknowledged. The study was based on data from selected healthcare facilities, which may limit generalizability to other populations and settings. Although the dataset included 1,601 patients, the sample size remains modest for training and evaluating multiple machine learning models. The absence of external validation further restricts the immediate applicability of the findings. Furthermore, the exclusive reliance on clinical features, although intended to improve feasibility, may constrain predictive performance, given that hematological markers like platelet count and hematocrit are recognized as significant predictors of dengue severity. The severity classification was based on clinical indicators recorded at a single time point, without accounting for the temporal progression of symptoms, which could provide additional predictive value. Furthermore, the exclusion of severe dengue cases (n = 1) limits applicability to the most critical patient group, although the analysis reflects real-world early-stage presentation scenarios. Future research should prioritize validating these models in larger, multi-center datasets, incorporating longitudinal clinical data, and exploring explainable artificial intelligence techniques to improve transparency and clinical usability. External validation and integration into real-world clinical workflows will be essential steps before deployment in routine practice.

Conclusions

In this study, machine learning models demonstrated the potential to accurately classify dengue cases according to warning sign status (DWS vs. DWWS) using routinely available clinical features. Ensemble-based approaches, particularly XGBoost and Random Forest, showed superior performance, highlighting their ability to capture complex, non-linear relationships among symptom-based predictors. These findings are consistent with prior research indicating that machine learning models can outperform conventional clinical criteria in identifying patients at risk of severe dengue. Importantly, the use of symptom severity scores without reliance on laboratory parameters enhances the feasibility of implementation in resource-limited settings. While the results are promising, external validation and evaluation in diverse populations are necessary before clinical adoption, particularly to ensure that the tools are effective across different demographic and clinical contexts. Overall, this study supports the potential of machine learning-based tools to assist early risk stratification and improve clinical decision-making in dengue management.

Abbreviations

AUC-ROC	Area Under the Receiver Operating Characteristic Curve
CRF	Case Report Form
DWWS	Dengue without Warning Signs
DWS	Dengue with Warning Signs
ML	Machine Learning

PR	Precision–Recall
RF	Random Forest
XGBoost	Extreme Gradient Boosting

Declarations

Ethics approval and consent to participate:

Ethical approval for this study was obtained from the Ethical Review Board (ERB) of the Nepal Health Research Council (Approval No. 45). All procedures were performed in accordance with the ethical standards of the Declaration of Helsinki and its subsequent amendments. Written informed consent was obtained from all adult participants and from parents or legal guardians for participants under 18 years of age prior to enrollment. To ensure confidentiality, participants were assigned unique identification numbers and all data were anonymized. Participation in the study was voluntary.

Consent for publication: Not applicable; no identifiable individual data are included.

Availability of data and materials:

The datasets and analysis code used in this study are available from the corresponding author upon reasonable request.

Competing interests:

CW, DPG, AKS, NA, ML, AP, SB, SU, PY, HB and BPG declare that they have no competing interests.

Funding:

The Virology Society Nepal provided a portion of the funding for this study.

Authors' contributions:

BPG conceptualized and supervised the study. HB and CW performed the data analysis. CW and NA drafted the manuscript. BPG, DPG, AKS, ML, AP, SB, PY, and SU contributed to the critical review and revision of the manuscript. All authors read and approved the final manuscript.

Acknowledgements:

The authors would like to thank the participating healthcare institutions and data collection teams for their support in this study.

References

1. World Health Organization. Dengue. 2025. https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue. Accessed 31 Aug 2025.
2. Haider N, Hasan MN, Onyango J, Billah M, Khan S, Papakonstantinou D, et al. Global dengue epidemic worsens with record 14 million cases and 9000 deaths reported in 2024. International Journal of Infectious Diseases. 2025;158:107940. https://doi.org/10.1016/j.ijid.2025.107940.
3. Lim A, Shearer FM, Sewalk K, Pigott DM, Clarke J, Ghouse A, et al. The overlapping global distribution of dengue, chikungunya, Zika and yellow fever. Nature Communications . 2025;16. https://doi.org/10.1038/S41467-025-58609-5,.
4. European Centre for Disease Prevention and Control. Dengue worldwide overview : Situation update, July 2025. https://www.ecdc.europa.eu/en/dengue-monthly. Accessed 31 Aug 2025.
5. World Health Organization South-East Asia. Dengue - SEARO. https://www.who.int/southeastasia/health-topics/dengue-and-severe-dengue. Accessed 31 Aug 2025.
6. Ooi E, Gubler DJ. Dengue in Southeast Asia: epidemiological characteristics and strategic challenges in disease prevention. Cad Saude Publica. 2009;25 SUPPL. 1:S115–24. https://doi.org/10.1590/S0102-311X2009001300011.
7. Epidemiology and Disease Control Division. SITUATION REPORT ON DENGUE IN NEPAL- 2023. 2023.
8. Epidemiology and Disease Control Division. SITUATION REPORT ON DENGUE IN NEPAL- 2024. 2024.
9. Chi M, Katuwal N, Shrestha A, Madhup SK, Tamrakar D, Shrestha R. Whole genome sequencing and phylogenetic analysis of dengue virus in Central Nepal from 2022 to 2023. BMC Global and Public Health. 2025;3. https://doi.org/10.1186/S44263-025-00135-Z.
10. Gauchan B, Mehanni S, Agrawal P, Pathak M, Dhungana S. Role of the general practitioner in improving rural healthcare access: A case from Nepal. Hum Resour Health. 2018;16. https://doi.org/10.1186/S12960-018-0287-7.
11. Gupta BP, Tuladhar R, Kurmi R, Manandhar K Das. Dengue periodic outbreaks and epidemiological trends in Nepal. Ann Clin Microbiol Antimicrob. 2018;17. https://doi.org/10.1186/S12941-018-0258-9.
12. World Health Organization. Dengue guidelines for diagnosis, treatment, prevention and control: new edition. World Health Organization; 2009.
13. Wiwanitkit V. Challenges in the diagnosis of dengue. Future Virol. 2020;15:485–7. https://doi.org/10.2217/FVL-2020-0058/ASSET//CMS/ASSET/D9FAD2AD-71EE-4FB8-9179-

27ACFE588BF1/FVL-2020-0058.FP.PNG.

14. Wong PF, Wong LP, AbuBakar S. Diagnosis of severe dengue: Challenges, needs and opportunities. J Infect Public Health. 2020;13:193–8. https://doi.org/10.1016/J.JIPH.2019.07.012.
15. Schaefer TJ, Panda PK, Wolford RW. Dengue Fever. BMJ Best Practice. 2024;:5–6.
16. Qasrawi R, Issa G, Thwib S, AbuGhoush R, Amro M, Ayyad R, et al. The role of machine learning in infectious disease early detection and prediction in the MENA region: A systematic review. Inform Med Unlocked. 2025;56:101651. https://doi.org/10.1016/J.IMU.2025.101651.
17. Liu M, Liu Y, Liu J. Machine Learning for Infectious Disease Risk Prediction: A Survey. ACM Comput Surv. 2025;57. https://doi.org/10.1145/3719663/SUPPL_FILE/CSUR-2023-0327-FILE002.PDF.
18. Andrade Girón DC, Marín Rodriguez WJ, Lioo-Jordan F de M, Ausejo Sánchez JL. Machine Learning and Deep Learning Models for Dengue Diagnosis Prediction: A Systematic Review. Informatics 2025, Vol 12, Page 15. 2025;12:15. https://doi.org/10.3390/INFORMATICS12010015.
19. Epidemiology and Disease Control Division. Prevention, Management and Control of Dengue in Nepal 2019. Kathmandu.
20. Omar ED, Mat H, Karim AZA, Sanaudi R, Ibrahim FH, Omar MA, et al. Comparative Analysis of Logistic Regression, Gradient Boosted Trees, SVM, and Random Forest Algorithms for Prediction of Acute Kidney Injury Requiring Dialysis After Cardiac Surgery. Int J Nephrol Renovasc Dis. 2024;17:197. https://doi.org/10.2147/IJNRD.S461028.
21. Madewell ZJ, Rodriguez DM, Thayer MB, Rivera-Amill V, Paz-Bailey G, Adams LE, et al. Machine learning for predicting severe dengue in Puerto Rico. Infect Dis Poverty. 2025;14. https://doi.org/10.1186/s40249-025-01273-0.
22. Castro PO, Martínez TE, Martínez-Rodríguez A, González RD, Guzmán TMG. Usefulness of warning signs in the management of dengue patients. Rev Cubana Med Trop. 2022;72.

Figures

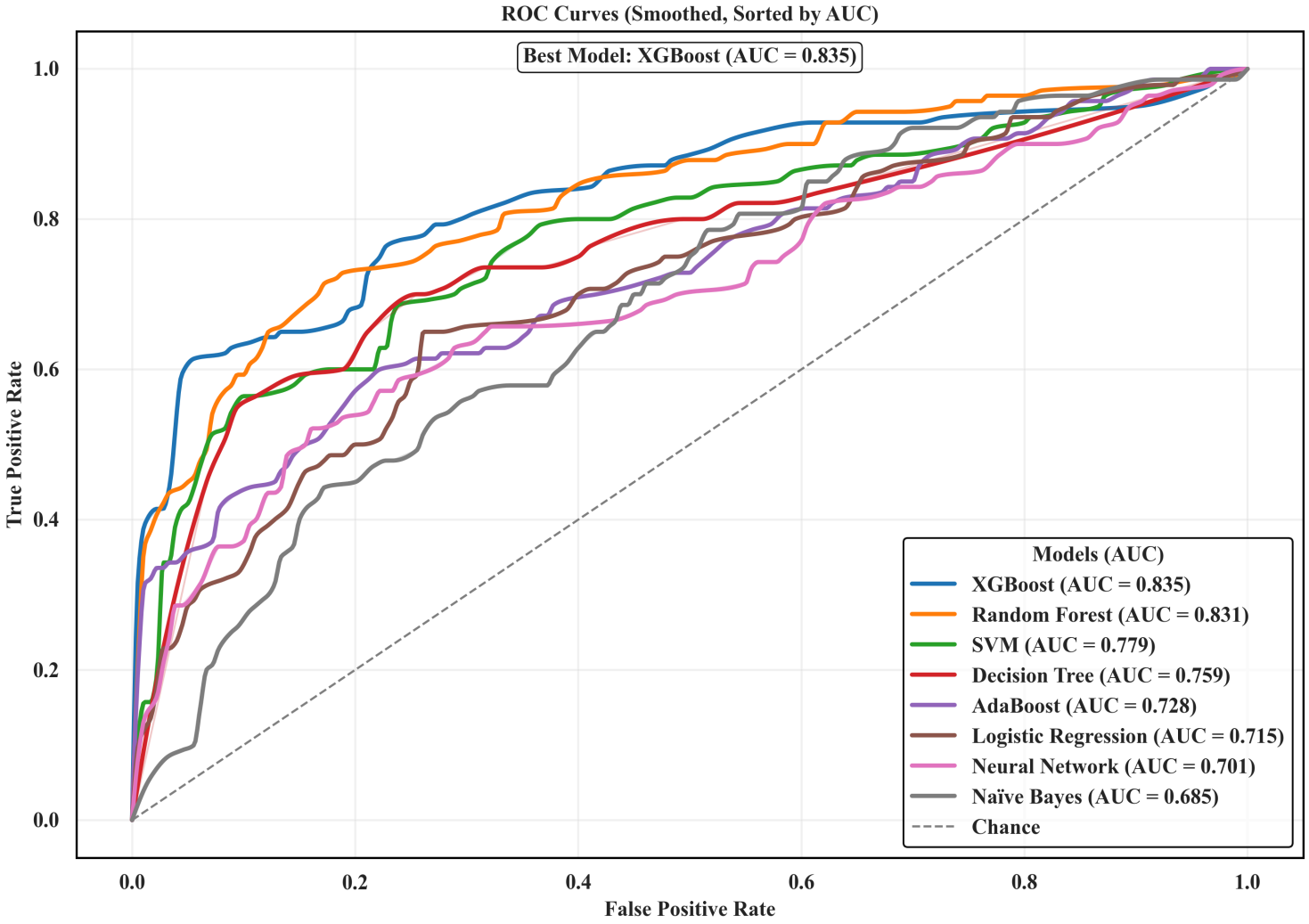

Figure 1

ROC and precision–recall curves of machine learning models

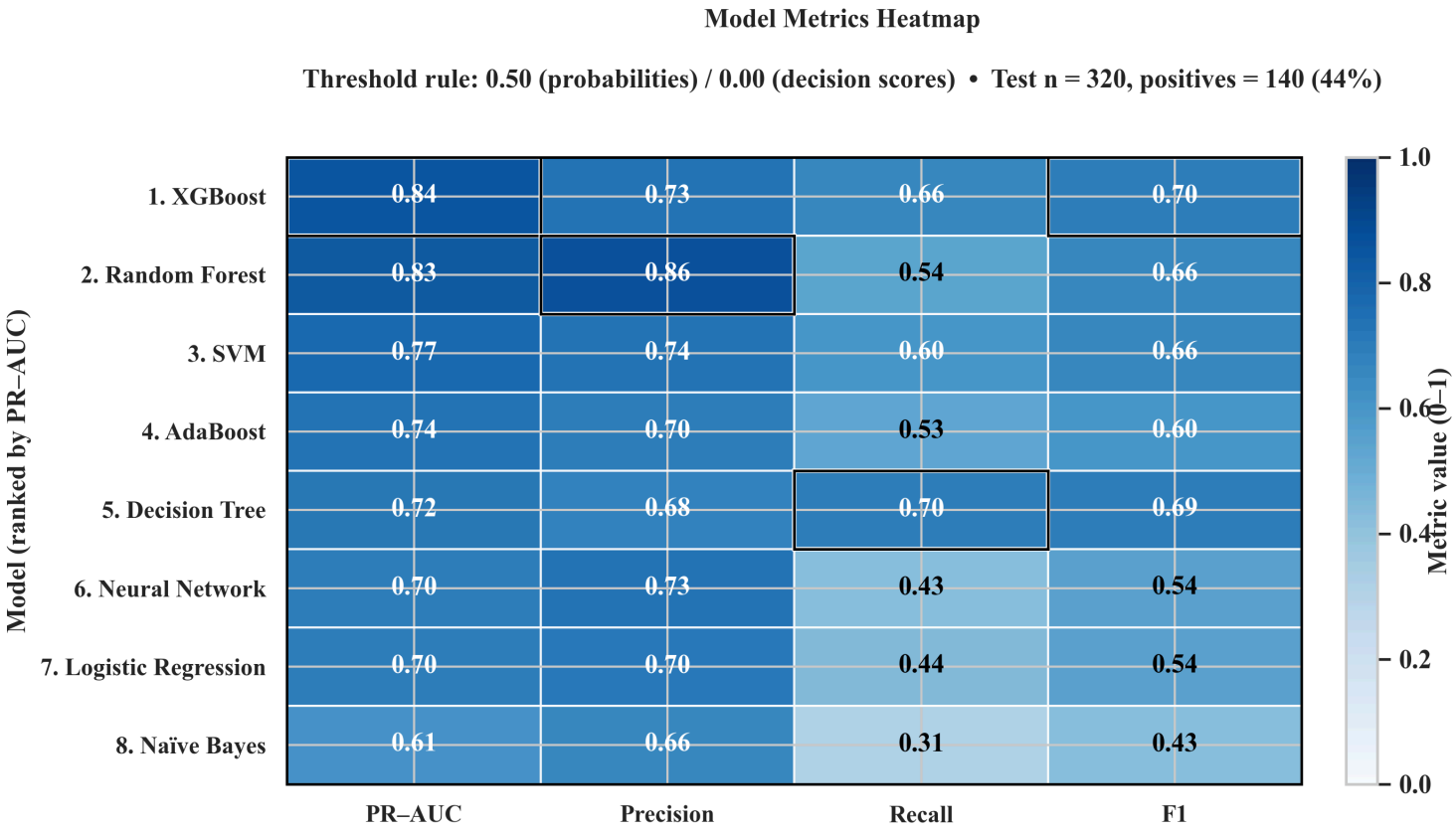

Figure 2

Precision–recall performance and threshold-based metrics of machine learning models

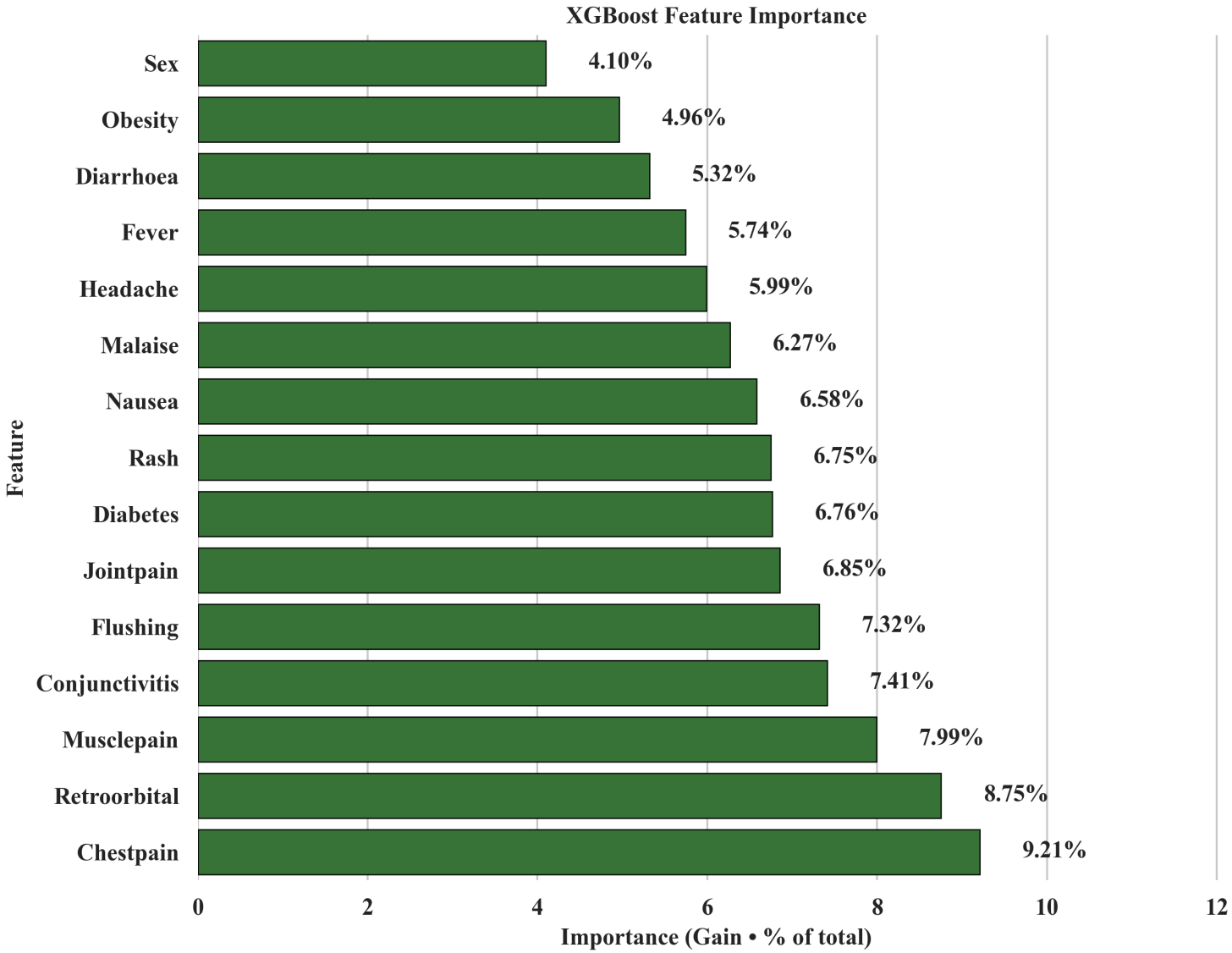

Figure 3

Feature importance of predictors based on the XGBoost model

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- CaseReportFormCRF.pdf