Journal Pre-proof

A Comprehensive Hematological Dataset for Dengue Incidence in Bangladesh

Md. Asraful Sharker Nirob , A K M Fazlul Kobir Siam , Prayma Bishshash , Md. Assaduzzaman , Md. Afzalul Haque , Arif Mahmud

PII: S2352-3409(25)00394-4
DOI: https://doi.org/10.1016/j.dib.2025.111664
Reference: DIB 111664

To appear in: *Data in Brief*

Received date: 19 February 2025
Revised date: 6 May 2025
Accepted date: 8 May 2025

Please cite this article as: Md. Asraful Sharker Nirob , A K M Fazlul Kobir Siam , Prayma Bishshash , Md. Assaduzzaman , Md. Afzalul Haque , Arif Mahmud , A Comprehensive Hematological Dataset for Dengue Incidence in Bangladesh, *Data in Brief* (2025), doi: https://doi.org/10.1016/j.dib.2025.111664

Data in Brief
Open access
Article template

ARTICLE INFORMATION

Article title

A Comprehensive Hematological Dataset for Dengue Incidence in Bangladesh

Authors

Md. Asraful Sharker Nirob[1], A K M Fazlul Kobir Siam[1], Prayma Bishshash[1], Md. Assaduzzaman[1,*], Md. Afzalul Haque[2], Arif Mahmud[1]

Affiliations

[1]Department of Computer Science and Engineering, Daffodil International University, Daffodil Smart City, Birulia 1216, Dhaka, Bangladesh

[2]Department of Medicine, Mymensingh Medical College Hospital, Mymensingh 2200, Dhaka, Bangladesh

Corresponding author's email address and Twitter handle

assaduzzaman.cse@diu.edu.bd

Keywords

Dengue Fever; Clinical Analysis, Statistical Exploration, Hematological Parameters, Public Health, Diagnostic Markers, Blood Indices, and Data Visualization.

Abstract

This study presents a comprehensive statistical and graphical analysis of hematological parameters in a cohort of 1523 dengue patients in a thorough search for the determination of the clinical features of dengue fever. The most important parameters were hemoglobin level, neutrophil and lymphocyte counts, and total platelet counts, whose statistical analysis has been conducted with the vision to clarify their role in dengue pathology. Statistical significance finds the key hematological parameters such as neutrophils, lymphocytes, monocytes, red blood cells, and RDW-CV to be strongly associated with dengue, and the same is borne out by their p-values. The gender, hemoglobin, MCH, and PDW are less strongly related, however. The consistency of results using T-tests as well as Z-tests strengthens the findings, highlighting significant markers while also pointing out how more intricate interactions would be uncovered using multivariate analysis. Subsequently, several data visualizations were constructed using box plots, violin plots, and pair plots. Hematological parameters in dengue positive and negative patients are somewhat disparate, as shown by box plots; positive patients presented with lower WBC and RBC counts and higher hemoglobin levels and monocyte percentages. Although females tend to have greater monocyte percentages and platelet counts and males tend to have overall greater hematocrit levels, violin plots do not indicate any gender differences in the majority of the blood parameters. The pair plot indicates that younger patients with lower hemoglobin are more likely to have positive dengue results, with some overlap between total WBC count and lymphocyte percentage. In conclusion, this database will be very useful and will provide a tremendous benefit to dengue research studies, public health initiatives, and better patient diagnosis and treatment methods.

SPECIFICATIONS TABLE

Subject	Computer Science.
Specific subject area	Public health, Epidemiology, Clinic Hematology, Prediction, Diagnosis & Treatment.
Type of data	Table, Dataset
Data collection	Data were collected from 1523 Dengue patients admitted to Jamalpur Medical College Hospital in Jamalpur, Bangladesh. Detailed clinical data were collected using standardized diagnosis tests for hemoglobin, white blood cell count, platelet count, and results. The demographic details on age and sex were obtained and recorded along with clinical data. This was done ethically, with due care taken that the data anonymized protective of privacy.
Data source location	Jamalpur Medical College Hospital in Jamalpur, Bangladesh.
Data accessibility	Repository name: Mendeley Data Data identification number: 10.17632/6fsrsk3mb8.2 Direct URL to data: https://data.mendeley.com/datasets/6fsrsk3mb8/2
Related research article	None.

1 VALUE OF THE DATA

- This database offers full hematological profiles of dengue patients, including critical parameters such as hemoglobin, total and differential white blood cell counts, and platelet counts that are helpful in dengue diagnosis and ascertainment of the severity of infection. Such detailed data are very valuable to the researcher and clinician for the refinement of diagnostic criteria and therapeutic protocols for dengue fever.
- The features of this dataset include, among others, results and blood indices such as MCV and PDW. his will allow researchers to create and validate predictive models that will be able to facilitate the early identification of dengue infection severe cases at recruitment points. This way, the potential for improvement in patient management increases through prompt and appropriate intervention based on disease severity predicted.
- The dataset therefore forms a very sound foundation for predictive model creation and validation that could classify and predict the severity of dengue infection. Machine learning and statistical modeling will hence provide an avenue through which researchers are able to develop algorithms

that interpret hematological parameter changes brought about by dengue infection, thereby providing valuable input at both the clinical decision-making and precision medicine levels.

- The data allows for inter-demographic group comparisons, and the variables included, like age, gender, and hematological profile, allow for conclusions about how the disease affects various populations. These possibly allow for molding public health responses and designing focused awareness campaigns in the middle of the seasonal outbreaks or public health vulnerability areas encounter.
- This therefore serves as a helpful handbook in an interdisciplinary research approach, combining the efforts of epidemiologists, clinicians, data scientists, and experts in public health. Furthermore, it creates an avenue for comparisons across regions with other comparable datasets from endemic areas, thus fostering a wider-based understanding of the effect of dengue on diverse populations and different health systems.

2. BACKGROUND

Dengue fever is a viral infection, of high public health significance in countries partially covering warm and tropical areas where Aedes mosquitoes are prevalent. In 2023, more than 80 countries in all WHO regions recorded the highest dengue cases on record. Sustained transmission and a record surge in dengue cases have already caused more than 6.5 million cases and more than 7300 dengue-related deaths since the start of 2023 [1]. These include not just the expansion of the distributions of vectors (primarily Aedes aegypti and Aedes albopictus mosquitoes) to dengue naïve countries previously [2] but also El Niño in 2023 and climate change effects, such as extreme temperature, torrential rainfall, and high humidity; compromised health systems due to the COVID-19 pandemic; political and economic instability, affecting various countries with complex humanitarian crises characterized by large population movements.

Dengue fever symptoms usually appear 4–10 days after being bitten by an infected Aedes mosquito. Dengue fever is associated with a sudden high temperature that may reach 104°F (40°C), and because of this, the illness has also been referred to as "breakbone fever [3]." This is usually accompanied by a bad headache, pain behind the eyes, and joint and muscle aches. Other common symptoms include skin rashes, swollen glands nausea, and vomiting. There is no specific treatment for Dengue [4]. Aimed at the symptomatic treatment of pain. Most dengue fever patients can be treated at home with pain medication. Acetaminophen —also called paracetamol—is commonly taken to treat pain. NSAIDs (anti-inflammatory [5] medications, whether or not steroid), such as aspirin and ibuprofen should be avoided, because they increase the risk of bleeding. Patients with severe dengue typically need to be hospitalized. A total dataset from 2013 to 2017 on dengue cases for the district in Selangor consisting of Gombak, Hulu Selangor, Hulu Langat, Klang, and Petaling was collected by Nurul Azam Mohd Salim & et al [6]. Dhiman Sarma et al [7]. Hosted a dataset including diagnostic reports, medical histories, and symptoms of dengue patients from real-time collected data in Chittagong Medical College Hospital and Dhaka Medical College Hospital in Bangladesh. Satya Ganesh Kakarla et al. [8] used the annual reports of the state surveillance unit of IDSP, Directorate of Health Services, Kerala for monthly dengue case data (2003–2017).

As hematological factors [9] are essential to diagnose and manage the disease, the collation of large datasets with comprehensive hematological profiles from cohort studies on dengue patients will be critical. Such databases can be instrumental in predictive model development, diagnostic accuracy, and our understanding of the disease mechanistically. An integrated approach of this kind is also required to Refine dengue fever research and clinical practice. We used a dataset obtained from Medical College Hospital, Jamalpur, Bangladesh containing different hematological parameters relevant to the diagnosis and management of Dengue fever. Critical parameters consist of demographic data (e.g., sex and age) along with important hematological measurements such as hemoglobin level [10], and total white blood cell count, neutrophil, lymphocyte, monocyte, eosinophil counts, and platelet counts. Hematocrit (HCT), MCV [11], mean corpuscular hemoglobin (MCH)/mean corpuscular hemoglobin concentration (MCHC) [12] and platelet distribution width (PDW) [13] should also be present in the data set. The dengue-positive test result gives you an idea about what type of dengue infection is occurring, which in turn aids in patient management and epidemiological research. The collection of this comprehensive data is important to enhance the accuracy of diagnosis and expand the current understanding of how various changes occurring in hematological parameters may influence the clinical outcomes at different phases of dengue infection.

3. DATA DESCRIPTION

This dataset consists of 19 hematological and clinical features collected from patients diagnosed with dengue and admitted to Jamalpur Medical College Hospital, Jamalpur, Bangladesh. While the full dataset includes 19 attributes, for the sake of clarity and improved readability, we focus on the 10 most essential features. These include demographic information such as age and gender, alongside key clinical markers such as hemoglobin levels, neutrophil count, platelet count, MPV, PDW, PCT, WBC count, and the diagnostic result. The "Result" attribute determines whether the test outcome is positive or negative. Table 1 presents these selected 10 features, offering a clear overview of the critical parameters involved in assessing the severity and health status of dengue infections.

Table 1: Overview of the dataset for dengue patients

Gender	Age	Hemoglobin	Neutrophils	Platelet	MPV	PDW	PCT	WBC	Result
Male	21	14.8	48	112000	10.7	15.4	0.12	5100	positive
Male	30	15	47	96000	10.6	15.8	0.121	4500	positive
Male	51	16.3	41	184000	10.4	16.4	0.13	6000	negative
Female	26	12.3	46	167000	8.1	17.1	0.11	5000	negative
Male	35	16.1	45	155000	10.52	12.34	0.15	4600	negative
Male	67	13.5	50	84000	9.7	14.6	0.101	4100	positive
Female	45	12.4	43	92000	10.3	13.6	0.183	4700	positive
Female	34	13.4	42	90000	8.9	13.6	0.111	4200	positive
Female	56	14.7	45	177000	10.23	16.1	0.141	5500	negative
Female	26	14.1	42	220000	10	17.6	0.17	6000	negative

An overview of the main clinical characteristics necessary for identifying and evaluating dengue fever is given in Table 2. A description, a typical reference range, and the kind of data (integer, float, or category) are all included in each attribute. Included are characteristics such as age, gender, and complete blood count (CBC) values, emphasizing their significance in monitoring the course of the disease. Variations in hemoglobin, neutrophils, and platelet count, for example, are extensively watched in dengue patients because they are linked to the severity of the disease. The table highlights how metrics like hematocrit (HCT) and red blood cell distribution width (RDW-CV) can indicate critical conditions like plasma leakage or thrombocytopenia, which are characteristic of severe dengue. It also includes reference values based on demographic categories (e.g., children, adults, male/female).

Table 2. Dataset feature explanation.

Attributes	Description	Reference Value	Types of values
Age	Age is typically recorded as the number of years	Years	Integer

Data in Brief
Open access
Article template
ELSEVIER

	since a patient's birth, playing a crucial role in assessing health patterns across age groups.		
Gender	Gender is recorded as a categorical variable, typically categorized as male or female.	Male or Female	Categorical
Hemoglobin(g/dl)	Hemoglobin levels in dengue patients can vary depending on disease severity.	M: 13-18, F: 11.5-16.5, C: 10-13, I: 8-10 g/dl	Float
Lymphocytes(%)	Lymphocyte levels can fluctuate in response to dengue infection.	C: 52-62, A: 20-50	Integer
Monocytes(%)	In dengue cases, monocyte percentages often fluctuate, with studies showing a tendency for either reduction or increase based on disease stage and severity.	C: 3-7, A: 2-10	Integer
Eosinophils(%)	In the context of dengue fever, eosinophil levels can provide insights into the immune system's status.	C:1-3, A: 1-6	Integer
RBC	Red blood cell (RBC) count is a vital parameter in the assessment of dengue fever as it provides insights into the patient's overall health and can indicate the severity of the disease.	M: 4.5-6.5, F: 3.8-5.5	Float
HCT(%)	A rising HCT can signify plasma leakage, a serious condition associated with dengue hemorrhagic fever, where fluid shifts from the vascular to the extravascular space.	M: 40-54, F: 37-47	Float
MVC(fl)	When it comes to dengue fever, MCV is essential for assessing the patient's overall hematologic profile and comprehending how the illness affects the shape and synthesis of red blood cells.	76-94	Float
MCH(pg)	Monitoring MCH levels in the setting of dengue fever might reveal important information about the patient's general health and anemia status.	27-32	Float
MCHC(g/dl)	In dengue patients, variations in MCHC can reflect changes in hydration status, hemoconcentration, or the severity of the disease.	29-34	Float
RDW-CV(%)	A useful hematological metric in dengue fever is the red cell distribution width-coefficient of variation (RDW-CV), which measures the variance in red blood cell (RBC) size.	10-16	Float
Total Platelet Count(/cumm)	Thrombocytopenia [14], a dramatic decrease in platelet count, is frequently seen in dengue fever and may be a warning of severe illness or upcoming consequences including shock or hemorrhagic symptoms.	150000-450000	Integer
MPV(fl)	MPV levels frequently rise in dengue patients during the acute stage of the disease, suggesting that the viral infection is causing an increased platelet turnover.	7-11	Float
PDW(%)	Since platelet size variability might reveal shifts in platelet production and activation, Platelet Distribution Width (PDW) is a crucial hematological metric in the setting of dengue fever.	10-18	Float
PCT(%)	Because PCT levels can drop dramatically as a result of increased platelet consumption and destruction during dengue infections, monitoring	0.1-0.2	Float

Data in Brief
Open access
Article template

	these levels is crucial.		
Total WBC Count(/cumm)	Since a notable decline in WBC counts may suggest a higher likelihood of severe dengue symptoms, monitoring these counts might provide information about the infection's intensity and possible side effects.	A: 4000-11000, C: 5000-15000, I: 6000-18000	Integer
Neutrophils(%)	Determining neutrophil percentages facilitates clinical decision-making and the assessment of illness progression.	A: 40-70, C: 50-80, I: 60-90	Integer
Result	A positive result indicates active infection, allowing timely management. A negative result suggests absence, but doesn't completely rule out the disease.	Positive or negative	Categorical

4. EXPERIMENTAL DESIGN, MATERIALS AND METHODS

Our dataset pertains to the hematological parameters of dengue patients treated at Jamalpur Medical College Hospital, Jamalpur, Bangladesh. It includes multiple variables such as Age, Sex, Hemoglobin Levels, WBC Count, Differential Count, RBC Panel, Platelet Count, and PDW. Additionally, the binary variable "Result" represents the final outcome of the diagnosis, indicating whether a patient tested positive or negative for dengue. Data preprocessing involves multiple steps, including cleaning the dataset to address missing values, normalizing numerical variables, encoding categorical variables, and imputing any missing data.

4.1. Statistical Analysis

Table 3 displays a statistical overview of important hematological parameters. With a mean age of 40.31 years, the dataset's age distribution shows that the patients' demographics are diverse. With a mean of 0.471, gender analysis shows an almost equal distribution. A hemoglobin level of 14.51 g/dl on average indicates sufficient oxygen-carrying capability. The mean neutrophil and lymphocyte counts, which are 44.48% and 43.56%, respectively, reflect immune responses. The mean platelet count is 173,127.17/cumm, which is an important metric for tracking the course of the illness. The significance of these factors in comprehending patient outcomes is shown by the high percentage of positive instances indicated by the overall test result average of 0.68. The necessary equations are given below:

Mean: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ (1)

Standard deviation (σ) = $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$ (2)

In equations 1 &2, n = total number of observations, i = index of summation, x_i = i^{th} data point in the dataset.

Table 3. Statistical Analysis Part-1.

Attributes	Mean	Standard Deviation	Min	25%	50%	75%	Max
Gender	0.471	0.499	0	0	0	1	1
Age	40.31	15.28	5	28	39	53	78
Hemoglobin(g/dl)	14.51	1.63	10.40	13.20	14.60	15.80	17.50

Data in Brief
Open access
Article template

Neutrophils(%)	44.48	4.34	29	42	45	48	60
Lymphocytes(%)	43.56	4.97	29	41	44	47	56
Monocytes(%)	3.27	1.21	2	2	3	4	9
Eosinophils(%)	3.08	1	1	2	3	4	9
RBC	4.76	0.72	4	4	5	5	7
HCT(%)	45.19	3.57	36.3	42.59	45.55	47.84	51.98
MCV(fl)	89.81	5.51	80	85.1	89.7	94.2	100
MCH(pg)	29.897	2.03	22.9	28.4	30	31.4	34
MCHC(g/dl)	32.05	1.55	27.08	30.9	32	33.2	35
RDW-CV(%)	13.63	1.43	11	12.4	13.8	14.7	21.33
Total Platelet Count(/cumm)	173127.17	64832.68	56000	114907	167385	228649.5	299803
MPV(fl)	9.59	0.85	7.5	8.9	9.66	10.28	11.23
PDW(%)	15.31	1.40	8.4	14.23	15.26	16.4	17.99
PCT(%)	0.289	5.99	0.00002	0.106	0.141	0.173	234
Total WBC Count(/cumm)	5483.02	1137.34	3500	4626	5397	6131.5	14900
Result	0.68	0.46	0	0	1	1	1

The overview of statistical analyses on several hematological and demographic characteristics in a dengue dataset is given in Table 4, which displays the findings of both the T-test and Z-test. Strong correlations with dengue are shown by the extremely low p-values of significant features such as neutrophils, lymphocytes, monocytes, red blood cells, and RDW-CV. Because the T-test and Z-test provide comparable p-values, the consistency of the results adds robustness and strengthens the validity of the findings. High p-values for characteristics like gender, hemoglobin, MCH, and PDW, however, imply that they might not be as important for differentiating dengue patients. Although this research successfully pinpoints important differentiators, it also points up a gap in its ability to capture intricate relationships, indicating the possible use of multivariate studies for more in-depth understanding.

$$\text{T-test statistic} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1{}^2}{n_1} + \frac{S_2{}^2}{n_2}}} \quad (3)$$

$$\text{Z-test statistic} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1{}^2}{n_1} + \frac{\sigma^2{}_1}{n_2}}} \quad (4)$$

Where (equation 3 & 4): $\bar{X}_1 - \bar{X}_2$= sample means of group 1 & 2, $(S_1{}^2, S_2{}^2)$ or $(\sigma_1{}^2, \sigma_2{}^2)$ = variance & n_1, n_2 = sample size.

Table 4. Statistical Analysis Part-2.

Attributes	T-test Statistic	T-Test (p-value)	Z-test Statistic	Z-Test (p-value)
Gender	0.4678	0.640	0.4678	0.640
Age	-3.538	0.0004	-3.538	0.0004
Hemoglobin(g/dl)	0.7722	0.4401	0.7722	0.4400
Neutrophils(%)	-6.7458	2.16e-11	-6.7458	1.52e-11
Lymphocytes(%)	4.4154	1.08e-05	4.4154	1.01e-05
Monocytes(%)	7.6555	3.40e-14	7.6555	1.93e-14

Data in Brief
Open access
Article template

Eosinophils(%)	5.7830	8.89e-09	5.7830	7.34e-09
RBC	6.8697	9.35e-12	6.8697	6.43e-12
HCT(%)	2.5143	0.0120	2.5143	0.0119
MCV(fl)	2.1505	0.0317	2.1505	0.0315
MCH(pg)	1.1449	0.2524	1.1449	0.2522
MCHC(g/dl)	-3.4389	0.0006	-3.4389	0.0006
RDW-CV(%)	7.6604	3.28e-14	7.6604	1.85e-14
Total Platelet Count(/cumm)	1.3986	0.1621	1.3986	0.1619
MPV(fl)	3.3188	0.0009	3.3188	0.0009
PDW(%)	0.0801	0.9362	0.0801	0.9362
PCT(%)	1.4481	0.1478	1.4481	0.1476
Total WBC Count(/cumm)	5.8264	6.90e-09	5.8264	5.67e-09

4.2. Data Visualization

Data visualization is the act of visually portraying our data to identify patterns and relationships [15]. Box plot, Violin Plot & Pair Plot were used to visualize data.

4.2.1. Data Visualization (Box plot)

A graphical technique for visualizing data distribution that helps with decision-making and insight-gathering is the box plot [16]. The box plots show minor variations in several hematological indicators between positive and negative dengue patients. In comparison to negative instances, positive patients have greater hemoglobin level fluctuation, higher monocyte percentages, and somewhat lower median WBC and RBC counts. MCH levels show modest variations, with positive patients showing somewhat higher levels, but RDW-CV, MPV, eosinophils, and lymphocytes show little variation between the two groups. Reduced WBC and RBC counts and elevated monocyte levels in positive cases are the main alterations overall.

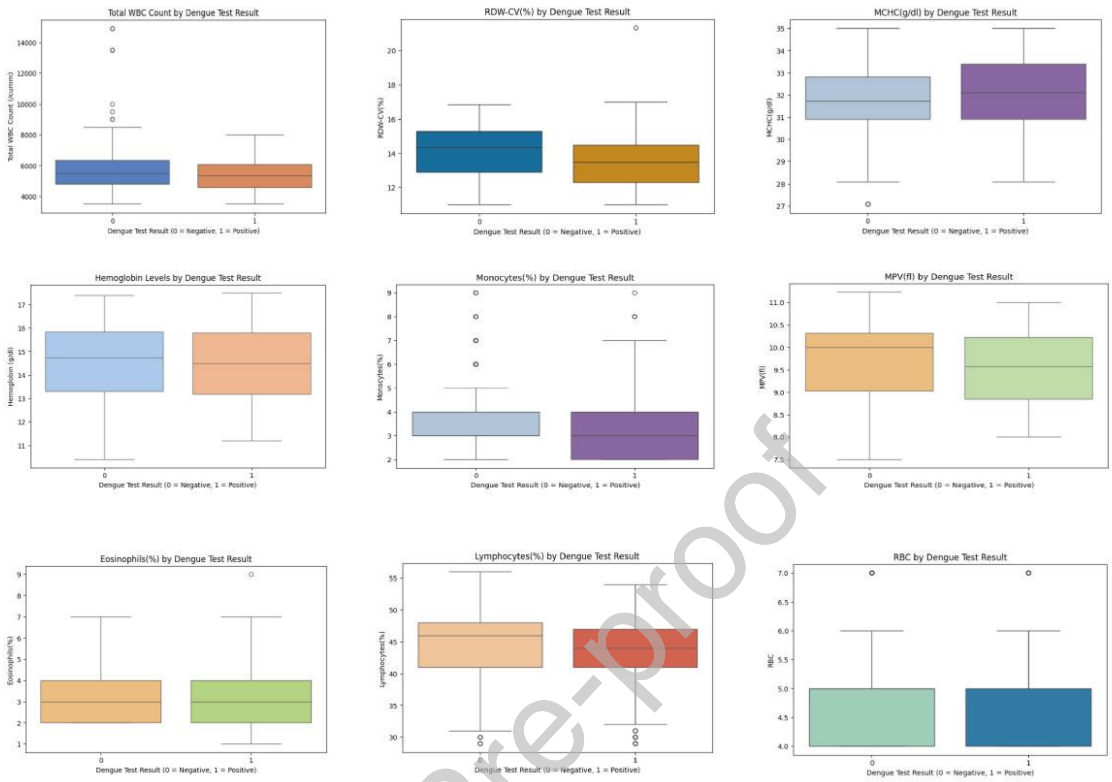

Fig. 1. Box plots of hematological parameters by Dengue test result (Positive vs. Negative).

4.2.2. Data Visualization (Violin Plot)

With a few significant exceptions, the violin plots, which depict the distribution of several blood parameters by gender (0 = Female, 1 = Male), often show comparable distributions across males and females. Males typically have somewhat greater hematocrit (HCT) values, which reflects normal physiological variances. The median platelet count is somewhat higher in females, while the range of monocyte percentages is wider in females, suggesting more variability. There are little to no gender-based variations in other metrics, including total WBC count, MPV, RDW-CV, MCH, neutrophil percentage, and MCV, indicating that these values have comparable distributions between the sexes.

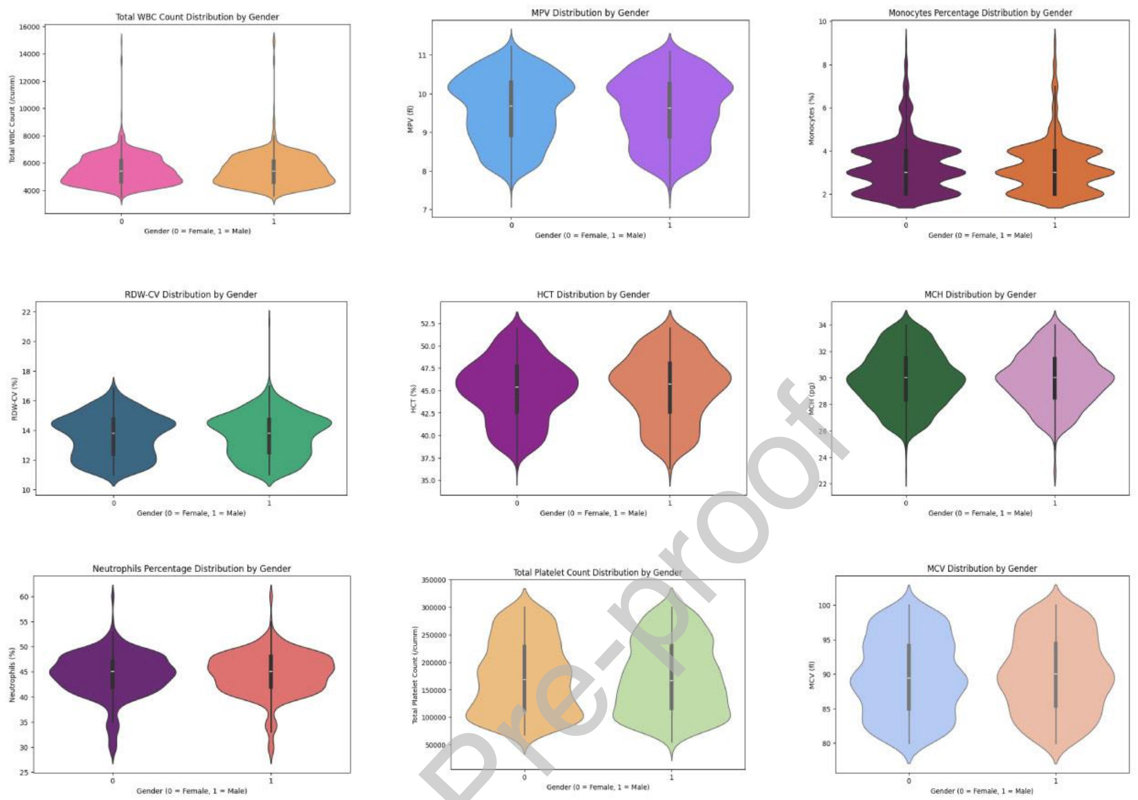

Fig. 2. Violin plots of hematological parameters by Gender (Female = 0, Male = 1).

4.2.3. Data Visualization (Pair Plot)

Relationships between hematological factors for both positive and negative dengue test results are displayed in the pair plot. Positive findings are more prevalent in younger ages (0–20 years), and positive results also tend to have lower hemoglobin levels and neutrophil percentages. However, factors such as lymphocyte percentage, MCV, and total WBC count exhibit significant overlap among test results, suggesting they are less useful in differentiating between positive and negative instances. Overall, certain factors lack distinct differences, while others exhibit modest trends with favorable outcomes.

Data in Brief
Open access
Article template
ELSEVIER

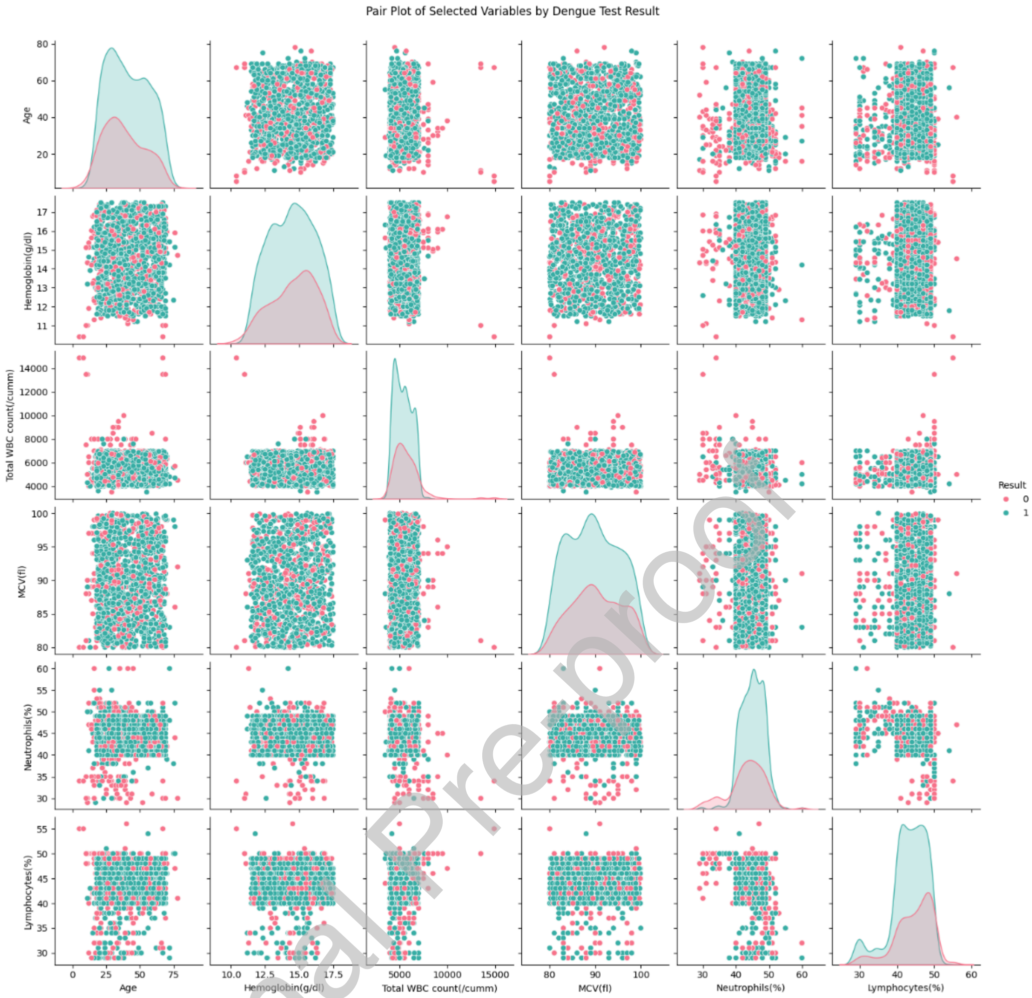

Fig. 3. Pair plot of selected hematological variables by Dengue test result.

Predictive analytics, clinical research, and public health initiatives might all benefit greatly from this dataset. This study uses 19 hematologic and demographic features, such as hemoglobin levels, platelet counts, and red and white blood cell parameters, to capture the minor hematologic changes linked to dengue infection. This lays the foundations for predictive modeling that will help with early diagnosis and disease severity assessment. In terms of epidemiology, this data could show trends within impacted groups, which might help with resource allocation during epidemics and guide focused treatments. Furthermore, by training algorithms to identify crucial hematological indicators suggestive of dengue, this dataset can be used to drive diagnostic automation through machine learning models, highlighting its applicability in both clinical and computational domains to improve research and disease management initiatives.

4.2.4. Data Visualization (Bar Plot)

The bar plot illustrates the gender distribution across dengue test results. It shows that male patients exhibit a slightly higher frequency of positive dengue results compared to females, although the distribution is relatively balanced. The number of negative cases is more evenly split between genders. This visualization provides insight into the gender-based differences in dengue test results,

Data in Brief
Open access
Article template

highlighting a potential trend in the prevalence of the disease among males, while also demonstrating that the disease affects both genders significantly.

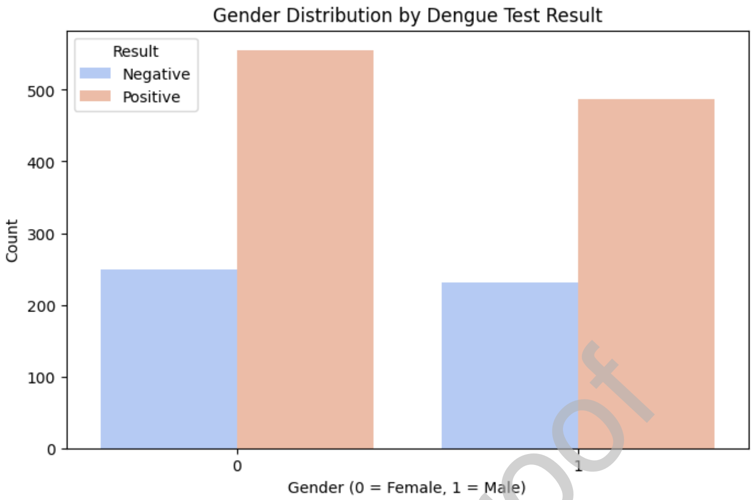

Fig. 4. Gender distribution by Dengue test result (Positive vs. Negative).

4.2.5. Data Visualization (Sankey Diagram)

The Sankey diagram visualizes the flow between different age groups and their corresponding dengue test results, categorized as either "positive" or "negative." The diagram clearly illustrates the distribution of test results across various age ranges.

From the flow, we observe that most cases in age groups such as 5-12, 13-20, and 21-28 are more likely to test negative. In contrast, age groups like 53-60, 69-76, and 77-80 show a higher proportion of positive results. The flow between positive and negative results appears more balanced in the middle age groups, with some age groups like 45-52 and 37-44 showing a relatively equal split.

Overall, the total number of negative test results is 481, while positive test results amount to 1.04k. The Sankey diagram helps in quickly identifying patterns in the distribution of test outcomes across age groups, providing valuable insights into age-based trends in dengue test results.

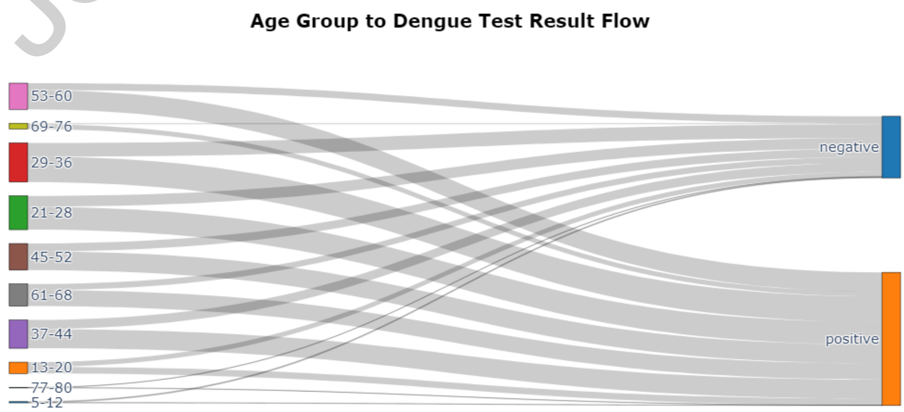

Fig. 5. Visual representation of the distribution of age groups across Dengue test results.

LIMITATIONS

This dataset, which comprises records from 1,523 dengue patients with 19 clinical and demographic features, has limitations in terms of generalizability because it was collected in a single site in Jamalpur, Bangladesh. Furthermore, climate-specific variables and variability, which might have a substantial impact on dengue severity and transmission, are not taken into account in the dataset. Lack of temporal and geographical variation in data collection may potentially limit the findings' generalizability to other climates or larger geographic areas.

ETHICS STATEMENT

Informed permission was provided by each participant or their legal guardian, and ethical guidelines were followed throughout the data collection procedure. The dataset was anonymized in order to protect patient privacy. The board of the Jamalpur Medical College Hospital, Jamalpur, Bangladesh, approved the facility on February 10, 2025. All data were used exclusively for the study in compliance with national and international ethical standards.

CRediT AUTHOR STATEMENT

Md. Asraful Sharker Nirob: Conceptualization, Methodology, writing – original draft, Data curation, **A K M Fazlul Kobir Siam:** Conceptualization, Visualization, Data curation, writing – original draft., **Prayma Bishshash**: Validation, Data curation, Writing. **Md. Assaduzzaman:** Writing – review & editing, **Md. Afzalul Haque:** Validation, Data curation, **Arif Mahmud:** Supervision

ACKNOWLEDGEMENTS

We would like to sincerely thank Assistant Registrar, Md. Jubayer Hossain, MBBS, BCS (Health), for his important support and help during the data-gathering procedure. His knowledge and commitment were crucial to the accomplishment of this study.

DECLARATION OF COMPETING INTERESTS

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

[1]H. T. Pham *et al.*, "Dengue Hemorrhagic Fever in Quang Nam Province (Vietnam) from 2020 to 2022—A Study on Serotypes Distribution and Immunology Factors," *Diagnostics*, vol. 14, no. 16, p. 1772, Aug. 2024, doi: https://doi.org/10.3390/diagnostics14161772.

[2]Fabien Vulu *et al.*, "Geographic expansion of the introduced Aedes albopictus and other native Aedes species in the Democratic Republic of the Congo," *Parasites & vectors*, vol. 17, no. 1, Jan. 2024, doi: https://doi.org/10.1186/s13071-024-06137-4.

[3]H. Ly, "Dengue fever in the Americas," *Virulence*, vol. 15, no. 1, Jul. 2024, doi: https://doi.org/10.1080/21505594.2024.2375551.

[4]G. N. Malavige and G. S. Ogg, "Molecular mechanisms in the pathogenesis of dengue infections," *Trends in Molecular Medicine*, vol. 30, no. 5, pp. 484–498, May 2024, doi: https://doi.org/10.1016/j.molmed.2024.03.006.

[5]A. Tsoupras, D. A. Gkika, I. Siadimas, I. Christodoulopoulos, P. Efthymiopoulos, and G. Z. Kyzas, "The Multifaceted Effects of Non-Steroidal and Non-Opioid Anti-Inflammatory and Analgesic Drugs on Platelets:

Data in Brief
Open access
Article template

Current Knowledge, Limitations, and Future Perspectives," *Pharmaceuticals*, vol. 17, no. 5, p. 627, May 2024, doi: https://doi.org/10.3390/ph17050627.

[6] N. A. M. Salim *et al.*, "Prediction of dengue outbreak in Selangor Malaysia using machine learning techniques," *Scientific Reports*, vol. 11, no. 1, Jan. 2021, doi: https://doi.org/10.1038/s41598-020-79193-2.

[7] D. Sarma, S. Hossain, T. Mittra, Md. A. M. Bhuiya, I. Saha, and R. Chakma, "Dengue Prediction using Machine Learning Algorithms," *IEEE Xplore*, Dec. 01, 2020. https://ieeexplore.ieee.org/document/9357035 (accessed Nov. 27, 2022).

[8] S. G. Kakarla *et al.*, "Weather integrated multiple machine learning models for prediction of dengue prevalence in India," *International Journal of Biometeorology*, pp. 1–13, Nov. 2022, doi: https://doi.org/10.1007/s00484-022-02405-z.

[9] F. R. F. G. Azin, R. P. Gonçalves, M. H. da S. Pitombeira, D. M. Lima, and I. C. Branco, "Dengue," *Revista Brasileira de Hematologia e Hemoterapia*, vol. 34, no. 1, pp. 36–41, 2011, doi: https://doi.org/10.5581/1516-8484.20120012.

[10] F. Zeb *et al.*, "Age, gender, and infectious status-wise assessments of hematological parameters among patients with dengue infection," *Heliyon*, vol. 10, no. 13, pp. e34053–e34053, Jul. 2024, doi: https://doi.org/10.1016/j.heliyon.2024.e34053.

[11]J. Guo *et al.*, "Chaperones Hsc70 and Hsp70 play distinct roles in the replication of bocaparvovirus minute virus of canines," *Molecular Microbiology*, vol. 121, no. 6, pp. 1127–1147, Apr. 2024, doi: https://doi.org/10.1111/mmi.15263.

[12]M. E. Day *et al.*, "Complete Blood Count Values Over Time in Young Children During the Dengue Virus Epidemic in the Dominican Republic From 2018 to 2020," *BioMed Research International*, vol. 2024, no. 1, Jan. 2024, doi: https://doi.org/10.1155/2024/3716786.

[13]M. Assaduzzaman, O. Islam, Md. A. S. Nirob, Md. M. H. Mim, and A. Mahmud, "A Benchmark Dataset for Analyzing Hematological Responses to Dengue Fever in Bangladesh," *Data in Brief*, p. 111030, Oct. 2024, doi: https://doi.org/10.1016/j.dib.2024.111030.

[14]J. Pierre *et al.*, "SEVERITY PREDICTION MARKERS IN DENGUE: A PROSPECTIVE COHORT STUDY USING MACHINE LEARNING APPROACH," *Biomarkers*, pp. 1–15, Nov. 2024, doi: https://doi.org/10.1080/1354750x.2024.2430997.

[15]Q. Zhang and N. Maram, "Multidimensional data visualization and synchronization for revealing hidden pandemic information," *Information Visualization*, Sep. 2024, doi: https://doi.org/10.1177/14738716241277559.

[16] T. G. Simon *et al.*, "Aspirin for Metabolic Dysfunction-Associated Steatotic Liver Disease Without Cirrhosis: A Randomized Clinical Trial," *JAMA*, vol. 331, no. 11, pp. 920–929, Mar. 2024, doi: https://doi.org/10.1001/jama.2024.1215.