Data in Brief 57 (2024) 111030

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

A benchmark dataset for analyzing hematological responses to dengue fever in Bangladesh

Md. Assaduzzaman[a,*], Oahidul Islam[b], Md. Asraful Sharker Nirob[a], Md. Minhajul Hayat Mim[a], Arif Mahmud[a]

[a] *Department of Computer Science and Engineering, Daffodil International University, Daffodil Smart City, Birulia, 1216 Dhaka, Bangladesh*
[b] *Department of Electrical & Electronic Engineering, Daffodil International University, Daffodil Smart City, Birulia, 1216 Dhaka, Bangladesh*

ARTICLE INFO

Article history:
Received 22 September 2024
Revised 4 October 2024
Accepted 7 October 2024
Available online 15 October 2024

Dataset link: Predictive Clinical Dataset for Dengue Fever Using Vital Signs and Blood Parameters (Original data)

Keywords:
Dengue fever
Hematological parameters
Predictive models
Clinical decision-making
Gender influence
Blood analysis

ABSTRACT

Dengue fever is an important public health problem in tropical and subtropical areas, and the understanding of its hematological changes is crucial for the improvement of diagnosis, treatment, and prognosis. This data set presents hematological parameters for the systematic record of patients suffering from dengue infection: age, sex, hemoglobin, WBC count, differential count, RBC panel, platelet count, and PDW. The dataset has in-depth records of patients admitted to Upazila Health Complex, Kalai, Jaipurhat, Bangladesh and thus offers an opportunity for further analysis of hematological changes produced by dengue infection. This dataset is valuable because of the potential contribution that these data will make to developing predictive models for disease severity and patient outcomes, enhancing clinical decision-making. It serves as a benchmark for the comparison of hematological responses across different demographics and geographical locations, adding value to the knowledge of dengue in the world. Moreover, the study has gone further to indicate how the characteristics related to blood may be affected by various

* Corresponding author.
E-mail address: assaduzzaman.cse@diu.edu.bd (M. Assaduzzaman).

https://doi.org/10.1016/j.dib.2024.111030

treatment regimens, hence offering better treatment protocols. The data preprocessing in this study involved cleaning, normalization, and encoding of the variables before proceeding to perform the statistical analysis. This showed a Chi-Square test for no significant association of sex with the diagnostic outcome, as given by the *p*-value of 0.277. On the other hand, the Z-test and T-test results indicated a significant difference in the hemoglobin levels concerning gender; the obtained p-values are 2.534×10^{-8} and 4.325×10^{-8}, respectively. These findings emphasize how gender influences hematological responses against dengue. In summary, this database will be a great help and will give a leading edge to the research studies of dengue, public health strategies, and improved diagnosis and treatment modalities for patients.

Specifications Table

Subject	Computer Science.
Specific subject area	Public health, Epidemiology, Clinical hematology, Prediction, diagnosis, and treatment of dengue fever
Type of data	Table, Dataset.
Data collection	The data were collected from patients diagnosed with dengue fever at Upazila Health Complex, Kalai, Jaipurhat, Bangladesh on 2nd November 2023. Ethical guidelines were strictly followed throughout data collection to ensure patient confidentiality and protection of personal information. Each patient's vital signs and blood profiles were recorded at diagnosis and follow-up visits. The dataset includes clinical measurements such as age, sex, hemoglobin level, white blood cell (WBC) count, differentiation rate, red blood cell (RBC) analysis, platelet count, and platelet distribution widely (PDW). To maintain data integrity, all patient records are anonymized, and any identifying information is removed. The dataset was cleaned and normalized for accuracy and use for further analysis. This comprehensive data collection system provides sufficient resources for the development of predictive models, for early diagnosis, and for understanding the clinical course of dengue fever in affected populations
Data source location	Upazila Health Complex, Kalai, Jaipurhat, Bangladesh
Data accessibility	Repository name: Mendaly Data Data identification number: 10.17632/xrsbyjs24t.1 Direct URL to data: https://data.mendeley.com/datasets/xrsbyjs24t/1
Related research article	none.

1. Value of the Data

- This dataset offers a comprehensive compilation of hematological parameters for dengue patients, encompassing hemoglobin levels, white blood cell count, differential count, red blood cell panel, platelet count, and platelet distribution width (PDW). An extensive dataset is essential for thorough analysis and comprehension of the hematological changes linked to dengue infection.
- The dataset is a highly significant resource for researchers seeking to develop biomarkers for the diagnosis and prognosis of dengue. Through the analysis of variations in hematological indicators, it is feasible to create predictive models that can evaluate the severity of diseases and forecast patient outcomes. This can ultimately improve the process of making clinical decisions.

- This dataset serves as a standard for comparing hematological reactions in dengue patients among different groups and geographical regions. This enhances the overall comprehension of the illness on a worldwide level and promotes the uniformity of clinical protocols.
- Scientists can utilize this information to assess the influence of different treatment regimens on the blood-related characteristics of individuals with dengue fever. The acquired knowledge can guide the creation and enhancement of treatment procedures, ultimately enhancing the quality of patient care and results.
- The dataset can be used by health professionals and policymakers to inform the development of strategies for managing and controlling dengue. The data can aid in the allocation of resources, the planning of healthcare services, and the formulation of guidelines based on evidence for the treatment of dengue.
- The dataset promotes multidisciplinary research by allowing the incorporation of hematology data with clinical, ecological, and sociodemographic variables. Thorough research like this can offer a complete understanding of infection with dengue and guide the development of diverse strategies for its control.

2. Background

Dengue fever poses a substantial public health obstacle, particularly in tropical and subtropical areas. The etiology of the disease is attributed to the virus that causes den which is primarily spread by Aedes mosquitoes. The World Health Organization, also known as the WHO, states that dengue is prevalent in more than 100 countries, with around 390 million cases of infection occurring each year, of which 96 million show clinical symptoms. The frequency of recorded dengue deaths is rather low, with the most frequently mentioned figure being approximately 20,000 deaths per year [1]. Dengue can manifest in a range of clinical presentations, from a low fever to more serious conditions including dengue hemorrhagic fever (DHF) and dengue shock syndrome (DSS) [2]. These severe forms are marked by the leaking of plasma, significant bleeding, and damage of organs. Dengue infection is characterized by hematological alterations, specifically thrombocytopenia (reduced platelet count) with leukopenia (reduced white blood cell count) [3], which are the primary laboratory observations. The evaluation of these blood-related factors is essential for the identification and treatment of dengue.

Thrombocytopenia is a strong indicator of serious cases of dengue and it is crucial to closely monitor platelet counts for effective patient care [4]. In addition, hemoconcentration, as seen by elevated hematocrit levels, is linked to plasma leakage and serves as an indicator of dengue hemorrhagic fever (DHF) [5]. Studies have examined white blood cell (WBC) counts, namely leukopenia and lymphocytosis, as potential indicators of illness severity. Although these hematological characteristics play a crucial role, there is a scarcity of comprehensive datasets that accurately capture complete hematological profiles of patients with dengue. Current research frequently concentrates on particular parameters or subsets of patients, resulting in incomplete knowledge of the comprehensive hematological alterations that occur during the progression of dengue illness. Luna et al. [6] provided data on dengue incidence in Brazil from 2014 to 2018. Dominguez-de-lo-Cruz et al. [7] provided dataset of Mexico related dengue virus serotypes and genotypes.

Due to the crucial importance of hematological parameters [8] in the therapy of dengue, it is necessary to have comprehensive datasets that offer thorough hematological profiles of patients with dengue. These databases have the potential to facilitate the creation of predictive models, boost the accuracy of diagnoses, and deepen our comprehension of the disease's etiology. The dataset offered in this study tries to address this deficiency by offering comprehensive hematological data from dengue patients who received treatment at Upazila Health Complex, Kalai, Jaipurhat, Bangladesh. The dataset contains characteristics such as gender, age, levels of hemoglobin, the count of white blood cells (WBC) [9], differential count, red blood cell (RBC) [10] panel, platelet count, and platelet distribution width (PDW) [11]. The extensive scope of this dataset enables a meticulous examination of the hematological alterations linked to dengue,

Table 1
Dataset overview of dengue patients' hematological parameters.

Age	Sex	Haemoglobin	WBC Count	Differential Count	RBC PANEL	Platelet Count	PDW	Final Output
43	Male	12.6	2200	1	1	62,000	11	1
45	Male	13.2	3000	0	1	17,000	17	1
50	Female	11	3300	1	1	19,000	16.3	1
57	Female	11.9	3500	1	0	29,000	14	1
51	Female	13	3100	0	1	30,000	14.5	1
61	Male	15	3300	1	1	34,000	20	1
6	Child	11	2300	1	0	69,000	12.5	1
21	Male	14	2500	1	1	77,000	13.3	1
29	Male	15	2400	1	1	78,000	14.5	1
31	Female	14.2	3700	0	1	82,000	15.6	1
37	Female	13.6	2300	1	0	81,000	16.7	1
39	Male	15.6	2200	1	1	86,000	17.3	1
44	Male	12.9	2900	0	1	99,000	18.2	1

facilitating the creation of models for prediction and enhancing diagnostic precision. This dataset can function as a standard for medical and epidemiological studies, facilitating the comparison of hematological responses in people with dengue across diverse demographics and geographies. Additionally, it can provide valuable insights for public health strategies and policies, hence enhancing the allocation of resources and planning of healthcare services for the management of dengue. The knowledge acquired from this collection of data will not only improve clinical procedures but also establish a basis for future studies in the realm of infectious illnesses and hematology.

3. Data Description

This dataset contains comprehensive hematological parameters obtained from dengue patients who received treatment at Upazila Health Complex, Kalai, Jaipurhat, Bangladesh. The dataset contains a range of crucial factors necessary for comprehending the hematological alterations linked to dengue infection. The variables encompassed in this set are the patient's age, sex, hemoglobin levels, white blood cell (WBC) count, differential count, red blood cell (RBC) panel, platelet count, and platelet distribution width (PDW). Furthermore, a conclusive output variable signifies the diagnostic or pertinent clinical results. This Table 1 provides a comprehensive overview of the key hematological parameters for patients diagnosed with dengue fever.

The following table presents a comprehensive explanation of each characteristic included in the dataset (Table 2).

4. Experimental Design, Materials and Methods

Our dataset is related to hematological parameters of dengue patients treated with many variables represented as Age, Sex, Hemoglobin levels, WBC Count, Differential Count, RBC PANEL, Platelet Count, and PDW. Also, the binary variable of the Final Output representing the results of the diagnosis will be done. Data preprocessing would involve several steps, including cleaning the dataset to handle the missing values, normalizing quantitative variables, encoding categorical variables, and imputing missing values.

4.1. Statistical analysis

The Chi-Square test [12] analyzed the association between Sex and Final Output. The test yielded a p-value of 0.277, indicating that there is no significant association between gender

Table 2
Feature description of dataset.

Feature	Description	Type	Unit/Value
Age	The patients' age, measured in years.	Numeric	Years
Sex	The gender of the patients, documented as Male or Female.	Categorical	Male. Female
Hemoglobin	Hemoglobin level is a measure of the blood's capacity to carry oxygen.	Numeric	g/dL
WBC Count	White Blood Cell count quantifies the concentration of white blood cells in a microliter of blood.	Numeric	Cells/μL
Differential Count	Distribution of several forms of white blood cells. The dataset has a binary indication here.	Binary	0,1
RBC PANEL	Different Red Blood Cell properties, compiled as a binary indication in the dataset.	Binary	0,1
Platelet Count	Platelet count, expressed as microliter of blood's number of plates.	Numeric	Cells/μL
PDW	Measure of the variety in platelet size: platelet distribution width.	Numeric	Percentage (%)
Final Output	Binary indicator for the diagnostic outcome or pertinent clinical result.	Binary	0,1

and diagnostic outcome. In addition to the Chi-Square test, a Z-test was performed to compare the means of various hematological parameters between male and female patients. The Z-test revealed a significant difference in hemoglobin levels, with a p-value of 2.534×10^{-8}, highlighting that gender affects hemoglobin levels. Furthermore, a T-test was conducted to compare the means of each quantitative variable between genders. The T-test for hemoglobin levels showed a significant difference with a p-value of 4.325×10^{-8} The T-test statistic is calculated as:

$$T = \frac{X_1 - X_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \tag{1}$$

Where X_1 and X_2 are the sample means, s_p is the pooled standard deviation, and n_1 and n_2 are the sample sizes for each group. This calculation confirms the significant difference in hemoglobin levels between male and female patients, further underscoring the impact of gender on this parameter. Overall, while the Chi-Square test did not find a significant association between sex and diagnostic outcome, both the Z-test and T-test revealed notable differences in hemoglobin levels based on gender, indicating that gender may influence this particular hematological parameter.

In addition to these tests, a one-way ANOVA (Analysis of Variance) was performed to compare hemoglobin levels across different diagnostic outcomes. The results are visualized in the following boxplot (Fig. 1), which shows the distribution of hemoglobin levels for each category of diagnostic outcome.

This box plot visualizes the distribution of hemoglobin levels across diagnostic outcome categories (final outcome). Boxplots show the interquartile range (IQR), median, and data point spread for each diagnostic outcome (0.0 and 1.0). A one-way ANOVA test was performed to determine if there was a significant difference in hemoglobin levels between the groups. Although the box plot shows some overlap in distribution, the ANOVA test showed no statistically significant differences between the groups.

In addition, the relationship between age and hemoglobin levels was explored to investigate whether age is a significant factor influencing hemoglobin. Fig. 2 displays the scatterplot for age versus hemoglobin levels.

This scatter plot shows the relationship between a patient's age and hemoglobin concentration. Each point represents an individual data point. The field highlights how hemoglobin levels vary with age in the data set. Based on this graph, no obvious correlation between age and

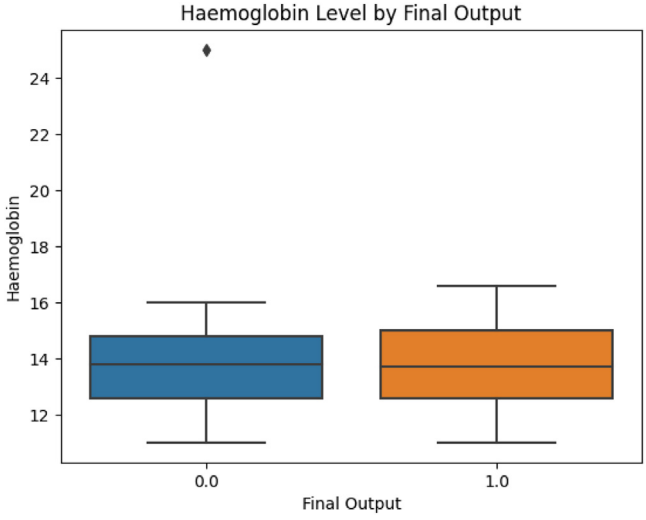

Fig. 1. Boxplot of hemoglobin levels by final output.

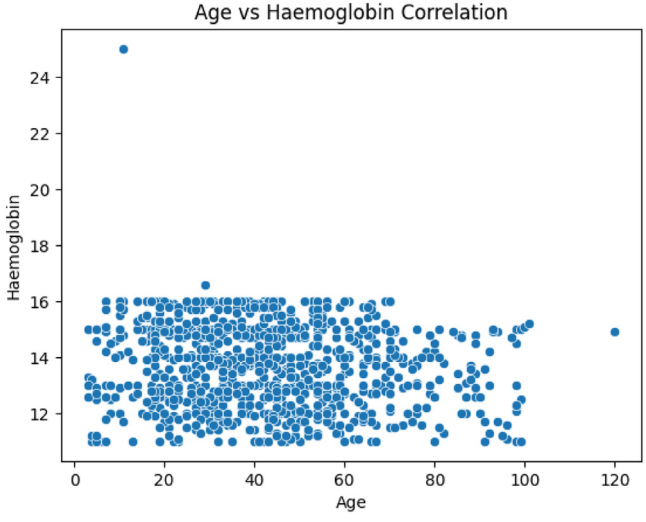

Fig. 2. Scatterplot of age vs hemoglobin levels.

hemoglobin concentration was observed. Fig. 3 presents a heatmap of the contingency table for sex versus final output, visualizing the distribution of diagnostic outcomes across sexes.

This heat map shows a contingency table of the last (0.0 and 1.0) corresponding to gender (male, female, and child). Darker colors indicate higher proportions of individuals in that category. The Chi-Square test was used to determine whether there was a significant association between gender and the final outcome of the study. The P-value of 0.277 indicated no significant association. This heat map provides a clear description of the distribution of different research outcomes across gender groups.

Finally, Fig. 4 shows the boxplots for quantitative variables by sex, illustrating how hematological parameters vary between males and females.

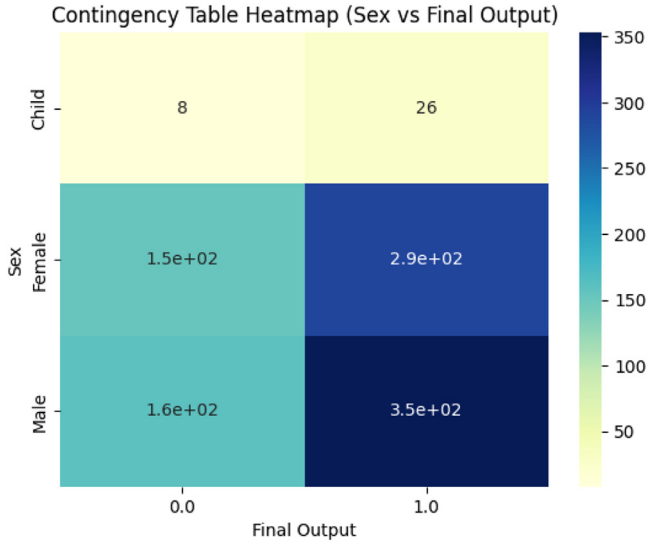

Fig. 3. Heatmap of the contingency table for sex vs final output.

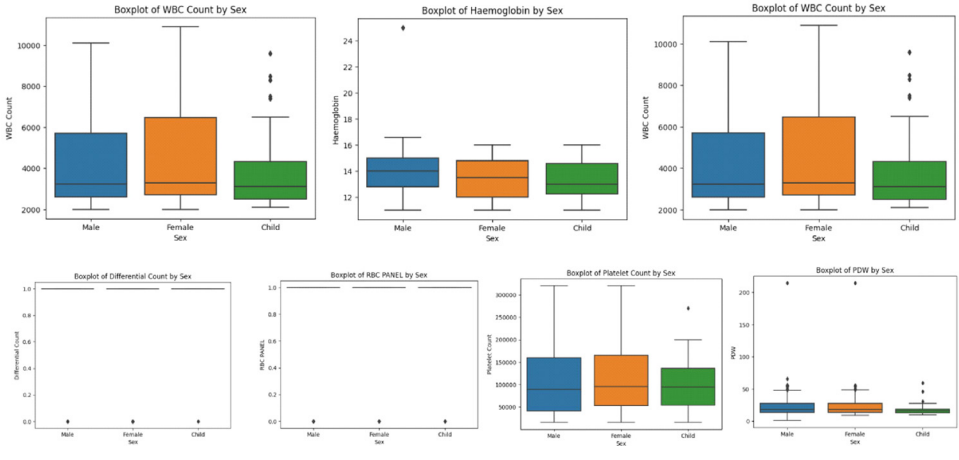

Fig. 4. Boxplots of hematological parameters by sex.

These boxplots display the distribution of various hematological parameters across male and female patients. The boxplots indicate the median, interquartile range (IQR), and potential outliers for each parameter. The Z-test and T-test were performed to compare the means of these parameters between the genders. For hemoglobin levels, the T-test revealed a significant difference between males and females, with a p-value of 4.325×10^{-8}, indicating that gender influences hemoglobin levels.

The results from the various statistical tests are summarized in Table 3, which lists the p-values for each test performed on the hematological parameters. These tests reveal significant gender-related differences in certain hematological parameters, such as hemoglobin, further highlighting the importance of considering gender as a variable in medical diagnostics.

Table 3
Statistical analysis of test result.

Test	Variable	*P*-value
Z-Test	Sex vs Final Output	0.277
Z-Test	Age	0.468
Z-Test	Hemoglobin	2.53e-08
Z-Test	WBC Count	0.190
Z-Test	Differential Count	0.172
Z-Test	RBC PANEL	0.075
Z-Test	Platelet Count	0.181
Z-Test	PWD	0.958
T-Test	Age	0.467
T-Test	Hemoglobin	4.35e-08
T-Test	WBC Count	0.192
T-Test	Differential Count	0.177
T-Test	RBC PNEL	0.071
T-Test	Platelet Count	0.180
T-Test	PWD	0.958

These analyses and visualizations collectively enhance our understanding of the hematological changes associated with dengue infection, contributing valuable insights to improve diagnostic and therapeutic strategies.

Limitations

The dataset offers valuable insights from Kalai, Joypurhat, but further validation in other regions could enhance generalizability. With 1003 entries, it captures a wide range of clinical responses to dengue fever, though larger datasets may provide more robust conclusions. While focused on key clinical parameters, future research could incorporate additional factors like comorbidities.

Ethics Statement

The data collection followed ethical guidelines, with informed consent obtained from all participants or their legal guardians. The dataset was anonymized to protect patient confidentiality. Approval was granted by the ethical review board of the **Upazila Health Complex, Kalai, Joypurhat, Bangladesh, on 2nd November 2023**. All data were used solely for research purposes following local and international ethical standards.

Credit Author Statement

Md Assaduzzaman: Conceptualization, Methodology, Data curation, **Oahidul Islam:** Conceptualization, Visualization, Data curation, Writing. **Arif Mahmud:** Supervision, Formal analysis, **Md. Asraful Sharker Nirob:** Writing – review & editing, **Md. Minhajul Hayat Mim:** Visualization, Data curation

Data Availability

Predictive Clinical Dataset for Dengue Fever Using Vital Signs and Blood Parameters (Original data) (Mendeley Data).

Acknowledgments

We would like to express our sincere gratitude to **Dr. Md. Faisal Nahid Pabitra**. Health Officer at the Upazila Health Complex, Kalai, Jaipurhat, Bangladesh, for his invaluable assistance in the data collection process.

Declaration of Competing Interest

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] J.D. Stanaway, et al., The global burden of dengue: an analysis from the global burden of disease study 2013, Lancet Infect Dis. 16 (6) (.2016) 712–723, doi:10.1016/S1473-3099(16)00026-8.
[2] M. Riaz, et al., Evaluation of clinical and laboratory characteristics of dengue viral infection and risk factors of dengue hemorrhagic fever: a multi-center retrospective analysis, BMC Infect. Dis. 24 (1) (2024) 500, doi:10.1186/s12879-024-09384-z.
[3] A. Dinkar, J. Singh, N. Kumar, K. Kumar, Santosh Kumar Singh, Arun Kumar Singh, Impact of secondary infections on dengue presentation: a cross-sectional study in a tertiary care hospital in Uttar Pradesh, India,", J. Infect. Public Health 16 (12) (2023) 1925–1932, doi:10.1016/j.jiph.2023.10.006.
[4] A. Wilder-Smith, E.-E. Ooi, O. Horstick, B. Wills, "Dengue,", The Lancet 393 (10169) (2019) 350–363, doi:10.1016/S0140-6736(18)32560-1.
[5] S.a.M. Kularatne, et al., Blood transfusion in severe dengue infection: a case series, J. Med. Case Rep. 17 (1) (2023) 17, doi:10.1186/s13256-022-03716-w.
[6] E. Luna, et al., Data on dengue incidence in South-eastern Brazil, 2014–2018,", Data Br. 29 (2020) 105266, doi:10.1016/j.dib.2020.105266.
[7] E. Domínguez-de-la-Cruz, et al., Dataset on the epidemiology and genetic diversification of dengue virus (DENV) serotypes and genotypes in Mexico, Data Br. 32 (2020) 106077, doi:10.1016/j.dib.2020.106077.
[8] G.H.P. Gonçalves, J.C. Souza Junior, H. da S. Pitz, A.R. Peruchi, F.S. Branco, Z.M.B. Hirano, Hematological and serum biochemistry data on southern brown howler monkeys (Alouatta guariba clamitans) in captivity in Brazil, J. Med. Primatol. 48 (6) (2019) 313–319, doi:10.1111/jmp.12431.
[9] N.M. Johannsen, et al., Effect of different doses of aerobic exercise on total white blood cell (WBC) and WBC subfraction number in postmenopausal women: results from DREW, PLoS ONE 7 (2) (2012) e31319, doi:10.1371/journal.pone.0031319.
[10] T. Aoki, A comprehensive review of our current understanding of red blood cell (RBC) glycoproteins, Membranes (Basel) 7 (4) (2017) 56, doi:10.3390/membranes7040056.
[11] Y. Zheng, L. Wang, Q. Shi, Mean platelet volume (MPV) and platelet distribution width (PDW) predict clinical outcome of acute ischemic stroke: a systematic review and meta-analysis, J. Clin. Neurosci. 101 (2022) 221–227, doi:10.1016/j.jocn.2022.05.019.
[12] M.L. McHugh, The chi-square test of independence, Biochem. Med. (Zagreb) 23 (2) (2013) 143–149, doi:10.11613/bm.2013.018.