

Supplementary Results

Tables

Supplementary Table 1. Definitions of the variants

Variant	features of the variant
V1	haemoglobinuria and haemolysis
V2	haemolysis and renal insufficiency
V3	cytopenia or MDS and haemolysis
V4	haemolysis and thrombosis
V5	haemolysis and Coombs negative
V6	cytopenia or MDS and thrombosis
V7	thrombosis
V8	bone marrow aplasia
V9	cytopenia or MDS and increased LDH
V10	anaemia and pulmonary hypertension
V11	haemolysis and incidental dysphagia
V12	haemolysis and iron deficiency
V13	haemolysis and erectile dysfunction
V14	haemoglobinuria
V15	PNH clone reassessment required

LDH: lactate dehydrogenase; MDS: myelodysplastic syndrome; PNH: paroxysmal nocturnal haemoglobinuria

Supplementary Table 2. Characteristic of the population of PNH patients used to train the algorithm (n=141)

Characteristic of confirmed PNH patient cohort used for algorithm training				
Symptom	Symptoms count	[%] of patients with symptom	Age average [years]	Age median [years]
Anaemia	124	87.94	46.85	43.00
Fatigue	123	87.23	47.28	45.00
Haemoglobinuria	101	71.63	47.35	43.00
Aplastic anaemia	95	67.38	43.14	36.00
Haemolysis	68	48.23	47.25	43.50
Pancytopenia	59	41.84	45.20	40.00
Episodic abdominal pain	57	40.43	44.07	40.00
Dyspnoea	56	39.72	48.36	46.50
Hypertension	55	39.01	56.35	61.00
Neutropenia	50	35.46	43.44	41.50
Headache	45	31.91	42.69	42.00
Myelodysplastic syndrome	40	28.37	53.55	54.00
Jaundice	38	26.95	43.92	37.00
Renal insufficiency	26	18.44	51.38	49.50
Haemolytic anaemia	16	11.35	56.19	50.00
Dark urine	14	9.93	43.36	46.50
Chronic kidney disease	12	8.51	44.33	38.50
Proteinuria	12	8.51	43.75	37.50
Decreased serum iron	12	8.51	49.08	47.50
Chest pain	11	7.80	55.00	60.00
Acute kidney injury	8	5.67	48.38	49.50
Coombs negative	8	5.67	41.00	31.00
Reduced haptoglobin level	7	4.96	46.86	42.00
Stroke	7	4.96	59.29	53.00
Pulmonary embolism	5	3.55	74.00	76.00
Glucosuria	4	2.84	65.25	66.00
Erythromelalgia	4	2.84	46.00	43.00
Myocardial infarction	4	2.84	72.50	74.00
Dysphagia	3	2.13	61.33	66.00
Pulmonary hypertension	2	1.42	53.00	53.00
Erythroid hyperplasia	2	1.42	56.50	56.50
Odynophagia	2	1.42	29.50	29.50
Mesenteric vessel thrombosis	1	0.71	27.00	27.00
Deep venous thrombosis	1	0.71	71.00	71.00
Impotence	1	0.71	45.00	45.00
Budd-Chiari syndrome	1	0.71	42.00	42.00
Hemosiderinuria	1	0.71	74.00	74.00

PNH: paroxysmal nocturnal haemoglobinuria

Supplementary Table 3. Full characteristics of the high-risk PNH and non-flagged patients.

	High-risk PNH cohort				Not flagged cohort				Age comparison	Symptom frequency comparison
Symptom	Symptoms count	% of patients with symptom	Age average	Age median	Symptoms count	% of patients with symptom	age average	age median	p. adj	p.adj
Fatigue	272	76.40	65.39	71.00	366 045	29.19	53.23	57.00	<0.001	<0.001
Anaemia	257	72.19	65.67	71.00	95 496	7.61	54.44	58.00	<0.001	<0.001
Hypertension	230	64.61	68.19	73.00	327 855	26.14	63.91	67.00	<0.001	<0.001
Myelodysplastic syndrome	175	49.16	73.26	74.00	3 027	0.24	68.06	72	0.003	<0.001
Dyspnoea	157	44.10	68.46	72.00	154 417	12.31	55.33	62.00	<0.001	<0.001
Headache	103	28.93	57.86	58.00	164 648	13.13	47.68	48.00	<0.001	<0.001
Episodic abdominal pain	74	20.79	62.76	71.50	162 446	12.95	47.22	49.00	<0.001	<0.001
Renal insufficiency	72	20.22	67.33	72.00	43 232	3.45	67.07	72.00	0.966	<0.001
Chest pain	62	17.42	68.40	74.00	55 811	4.45	60.01	64.00	<0.001	<0.001
Haemolysis	55	15.45	59.73	67.00	6 792	0.54	59.89	66.00	0.899	<0.001
Decreased serum iron	54	15.17	62.98	69.50	19 569	1.56	51.74	53.00	0.003	<0.001
Chronic kidney disease	47	13.20	68.00	74.00	47 294	3.77	66.64	71.00	0.481	<0.001
Stroke	46	12.92	62.78	65.00	39 022	3.11	69.66	72.00	0.037	<0.001
Jaundice	43	12.08	65.09	71.00	23 237	1.85	32.78	20.00	<0.001	<0.001
Portal vein thrombosis	41	11.52	56.76	57.00	604	0.05	57.40	58.00	0.898	<0.001
Myocardial infarction	37	10.39	71.41	75.00	50 816	4.05	69.03	71.00	0.360	<0.001
Aplastic anaemia	31	8.71	59.61	66.00	419	0.03	46.51	49.00	0.010	<0.001
Coombs negative	27	7.58	62.59	69.00	126	0.01	51.03	53.00	0.040	<0.001
Proteinuria	25	7.02	58.64	64.00	27 605	2.20	50.40	54.00	0.167	<0.001
Cerebral venous sinus thrombosis	24	6.74	49.50	49.50	310	0.02	45.47	48.00	0.663	<0.001
Haemolytic anaemia	23	6.46	59.70	58.00	684	0.05	53.76	59.00	0.414	<0.001
Reduced haptoglobin level	21	5.90	61.24	68.00	220	0.02	55.53	62.00	0.385	<0.001
Acute kidney injury	19	5.34	62.95	70.00	8 983	0.72	68.20	71.00	0.430	<0.001
Pulmonary hypertension	18	5.06	62.44	65.50	8 334	0.66	68.64	72.00	0.201	<0.001
Pulmonary embolism	17	4.78	58.71	61.00	12 527	1.00	66.63	69.00	0.120	<0.001
Dysphagia	11	3.09	67.55	69.00	10 432	0.83	53.61	58.00	0.138	<0.001
Deep venous thrombosis	10	2.81	67.00	75.50	8 168	0.65	64.11	67.00	0.481	<0.001
Thromboembolism	9	2.53	55.67	56.00	20 615	1.64	58.64	61.00	0.635	0.304
Glucosuria	9	2.53	65.67	69.00	15 922	1.27	51.73	54.00	0.105	0.073
Dark urine	9	2.53	63.67	63.00	3 701	0.30	62.90	67.00	0.899	<0.001
Haemoglobinuria	8	2.25	55.63	68.50	172	0.01	47.92	47.00	0.414	<0.001
Budd-Chiari syndrome	6	1.69	32.33	27.50	69	0.01	42.26	43.00	0.466	<0.001
Arterial thrombosis	5	1.40	38.80	47.00	289	0.02	65.27	68.00	0.016	<0.001
Erythromelalgia	4	1.12	76.25	77.00	7 711	0.61	54.50	58.00	0.078	0.395

Hemosiderinuria	4	1.12	63.00	70.00	294	0.02	64.53	69.00	0.899	<0.001
Erythroid hyperplasia	2	0.56	81.00	81.00	48	<0.01	63.63	66.00	0.261	<0.001
Increased circulating LDH	2	0.56	70.50	70.50	58	<0.01	57.60	61.00	0.532	<0.001
Odynophagia	1	0.28	56.00	56.00	2 359	0.19	48.82	49.00	0.850	1.000
Mesenteric vessel thrombosis	1	0.28	25.00	25.00	29	<0.01	64.69	66.00	0.201	<0.001
Neutropenia	ND	ND	ND	ND	13 842	1.10	52.75	60.00	NA	NA
Myelodysplastic syndrome	ND	ND	ND	ND	3 027	0.24	68.06	72.00	NA	NA
Pancytopenia	ND	ND	ND	ND	2 630	0.21	56.69	63.00	NA	NA
Impotence	ND	ND	ND	ND	433	0.03	55.28	56.00	NA	NA
Renal vein thrombosis	ND	ND	ND	ND	65	0.01	57.02	56.00	NA	NA
Oesophageal spasms	ND	ND	ND	ND	62	<0.01	67.60	69.00	NA	NA
Cerebral artery thrombosis	ND	ND	ND	ND	6	<0.01	52.67	54.00	NA	NA

LDH: lactate dehydrogenase; PNH: paroxysmal nocturnal haemoglobinuria

Supplementary Table 4. Comparison of characteristics between tested and non-tested patients identified as high-risk PNH by the algorithm.

Feature	Test	p.adj
Erythromelalgia	Fisher	0,73
Reticulocytosis	Chi-squared	0,81
Episodic abdominal pain	Chi-squared	0,81
Dark urine	Fisher	0,87
Budd-Chiari syndrome	Fisher	1
Deep venous thrombosis	Chi-squared	1
Unconjugated hyperbilirubinemia	Fisher	1
Cerebral venous sinus thrombosis	Chi-squared	1
Chest pain	Chi-squared	1
Coombs negative	Chi-squared	1
Odynophagia	Fisher	1
Anaemia	Chi-squared	1
Myocardial infarction	Chi-squared	1
Renal insufficiency	Chi-squared	1
Thromboembolism	Fisher	1
Pancytopenia	Chi-squared	1
PNH clone	Chi-squared	1
Jaundice	Chi-squared	1
Decreased serum iron	Chi-squared	1
Myelodysplastic syndrome	Chi-squared	1
Portal vein thrombosis	Chi-squared	1
Headache	Chi-squared	1
Pulmonary hypertension	Chi-squared	1
Arterial thrombosis	Fisher	1
Thrombocytopenia	Chi-squared	1
Increased circulating LDH	Chi-squared	1
Number of visits	Wilcoxon	1
Acute kidney injury	Chi-squared	1
Haemoglobinuria	Fisher	1
Haemolytic anaemia	Chi-squared	1
Haemolysis	Chi-squared	1
Proteinuria	Chi-squared	1
Leukopenia	Chi-squared	1
Chronic kidney disease	Chi-squared	1
Stroke	Chi-squared	1
Pulmonary embolism	Chi-squared	1

Hypertension	Chi-squared	1
Fatigue	Chi-squared	1
Dyspnoea	Chi-squared	1
Patient age	Wilcoxon	1
Reduced haptoglobin level	Chi-squared	1
Aplastic anemia	Chi-squared	1
Increased blood urea nitrogen	Chi-squared	1
Dysphagia	Fisher	1
Erythroid hyperplasia	Fisher	1
Glucosuria	Fisher	1
Hemosiderinuria	Fisher	1
Mesenteric vessel thrombosis	Fisher	1

LDH: lactate dehydrogenase; PNH: paroxysmal nocturnal haemoglobinuria

Supplementary Table 5. Relative importance of patient features discriminating tested and non-tested high-risk PNH patients, calculated using the MCFS-ID algorithm.

Position in the ranking	Feature Name	Number of projections	Relative Importance value (RI)
1	Reticulocytosis	92	<0.01
2	Cerebral venous sinus thrombosis	91	<0.01
3	Coombs negative	103	<0.01
4	Number of visits	102	<0.01
5	Deep venous thrombosis	93	<0.01
6	PNH clone	96	<0.01
7	Patient age	101	<0.01
8	Acute kidney injury	109	<0.01
9	Jaundice	123	<0.01
10	Renal insufficiency	122	<0.01
11	Dark urine	95	<0.01
12	Stroke	98	<0.01
13	Haemolytic anaemia	87	<0.01
14	Dyspnoea	96	<0.01
15	Aplastic anaemia	117	<0.01
16	Myelodysplastic syndrome	107	<0.01
17	Chest pain	105	<0.01
18	Episodic abdominal pain	106	<0.01
19	Anaemia	117	<0.01
20	Haemolysis	95	<0.01
21	Pancytopenia	106	<0.01
22	Reduced haptoglobin level	109	<0.01
23	Thrombocytopenia	93	<0.01
24	Leukopenia	104	<0.01
25	Myocardial infarction	99	<0.01
26	Pulmonary hypertension	97	<0.01
27	Hypertension	106	<0.01
28	Fatigue	113	<0.01
29	Decreased serum iron	115	<0.01
30	Headache	102	<0.01
31	Chronic kidney disease	103	<0.01
32	Pulmonary embolism	97	<0.01

33	Increased circulating LDH	83	<0.01
34	Erythromelalgia	102	<0.01
35	Increased blood urea nitrogen	94	<0.01
36	Proteinuria	108	<0.01
37	Portal vein thrombosis	98	<0.01
38	Dysphagia	93	<0.01
39	Thromboembolism	117	<0.01
40	Glucosuria	108	<0.01
41	Unconjugated hyperbilirubinemia	101	<0.01
42	Budd-Chiari syndrome	111	<0.01
43	Haemoglobinuria	103	<0.01
44	Arterial thrombosis	98	<0.01
45	Erythroid hyperplasia	105	<0.01
46	Hemosiderinuria	88	<0.01
47	Impotence	97	<0.01
48	Mesenteric vessel thrombosis	113	<0.01
49	Neutropenia	101	<0.01
50	Odynophagia	105	<0.01

LDH: lactate dehydrogenase; PNH: paroxysmal nocturnal haemoglobinuria

Supplementary Table 6. Performance metrics across incremental subsets of top-ranked features using multiple machine learning classifiers. Classification results include accuracy and weighted accuracy for: J48 (C4.5 Decision Tree), RF (Random Forest), NB (Naive Bayes), SVM (Support Vector Machine), k-NN (k-Nearest Neighbors), Logistic (Logistic Regression), and RIPPER (Repeated Incremental Pruning to Produce Error Reduction), calculated using MCFS-ID.

Number of top features	Algorithm	Accuracy (acc)	Weighted Accuracy (wacc)
1	j48	0,63	0,49
1	rf	0,64	0,50
1	nb	0,65	0,50
1	svm	0,63	0,49
1	knn	0,62	0,49
1	logistic	0,62	0,50
1	ripper	0,63	0,49
2	j48	0,62	0,48
2	rf	0,62	0,52
2	nb	0,65	0,50
2	svm	0,61	0,47
2	knn	0,60	0,48
2	logistic	0,62	0,49
2	ripper	0,61	0,48
3	j48	0,62	0,48
3	rf	0,64	0,51
3	nb	0,65	0,50
3	svm	0,61	0,47
3	knn	0,64	0,50
3	logistic	0,61	0,49
3	ripper	0,61	0,47
4	j48	0,65	0,52
4	rf	0,61	0,56
4	nb	0,65	0,50
4	svm	0,63	0,49

4	knn	0,53	0,49
4	logistic	0,62	0,50
4	ripper	0,62	0,48
5	j48	0,61	0,48
5	rf	0,59	0,55
5	nb	0,65	0,50
5	svm	0,63	0,50
5	knn	0,56	0,52
5	logistic	0,62	0,49
5	ripper	0,64	0,50
6	j48	0,64	0,50
6	rf	0,56	0,52
6	nb	0,65	0,50
6	svm	0,62	0,49
6	knn	0,55	0,51
6	logistic	0,63	0,51
6	ripper	0,60	0,48
8	j48	0,62	0,49
8	rf	0,60	0,55
8	nb	0,65	0,50
8	svm	0,63	0,50
8	knn	0,58	0,54
8	logistic	0,61	0,48
8	ripper	0,62	0,48

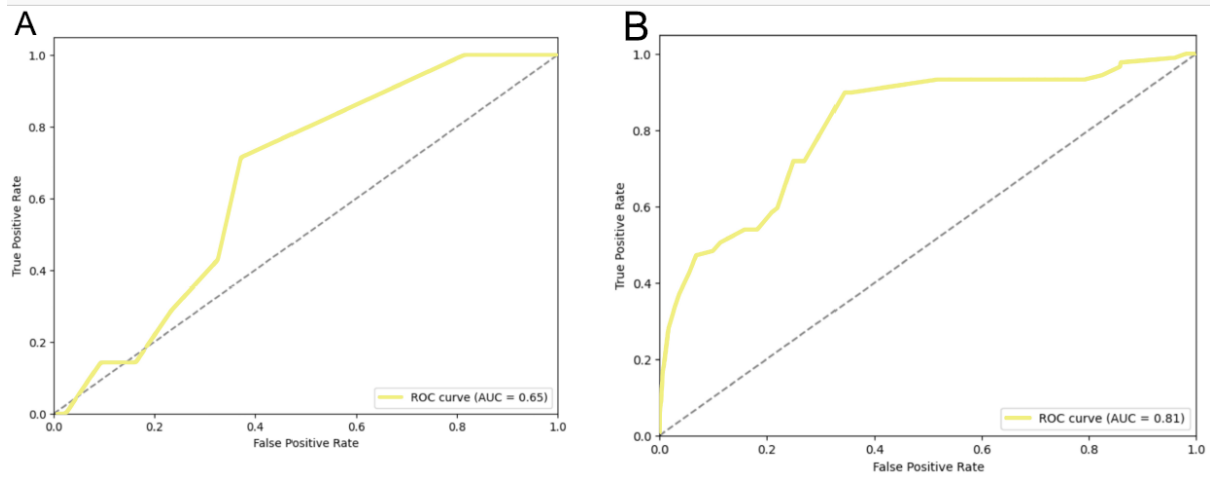
j48 - decision trees; knn - k-nearest neighbors; logistic - logistic regression; nb - Naïve Bayes; rf - random forest; ripper – ripper; svm - support vector machines

Supplementary Table 7. Clinical history of patients from a non-flagged cohort, who were subsequently diagnosed with PNH during the follow up analysis

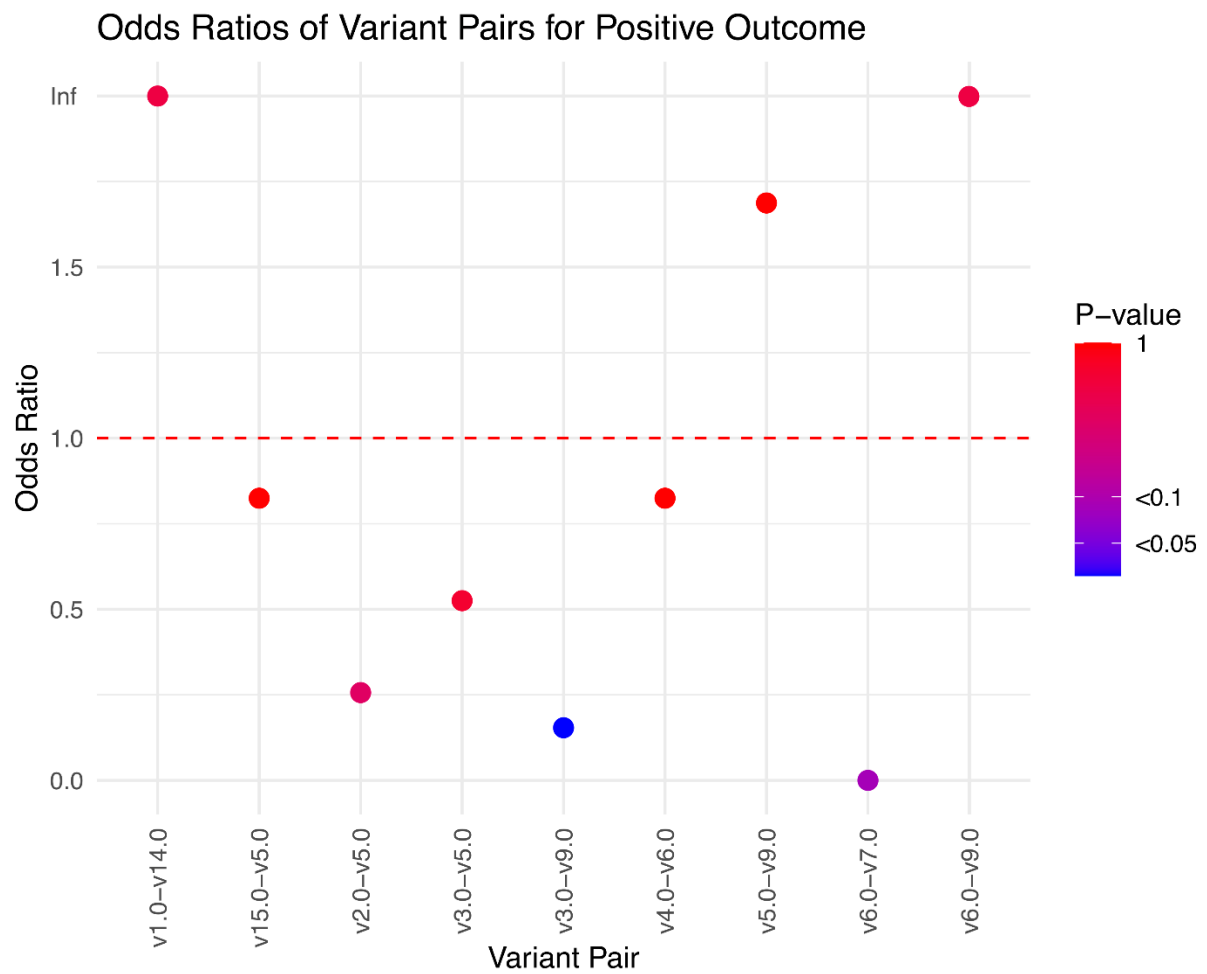
Patient	Diagnostic history	Date of PNH diagnosis
1	diagnostic work-up began in October 2024, with PNH testing ordered in November 2024 leading to a confirmed diagnosis in 2025	2025
2	the patient was referred in September 2024 for pancytopenia and suspected myelodysplastic syndrome (MDS), with a PNH diagnosis confirmed on December 4, 2024	4 th Dec 2024
3	The patient was admitted in November 2024 for pancytopenia and was diagnosed with MDS with a concurrent PNH clone >1% following the initial evaluation	Nov 2024
4	patient had no relevant haematological history during the screening period (only information available in EHR was completed cataract treatment in October 2024); PNH was diagnosed in November 2025 after referral to haematology in 2025 for pancytopenia, with testing initiated at the first visit	Nov 2025

EHR: electronic health record; MDS: myelodysplastic syndrome; PNH: paroxysmal nocturnal haemoglobinuria

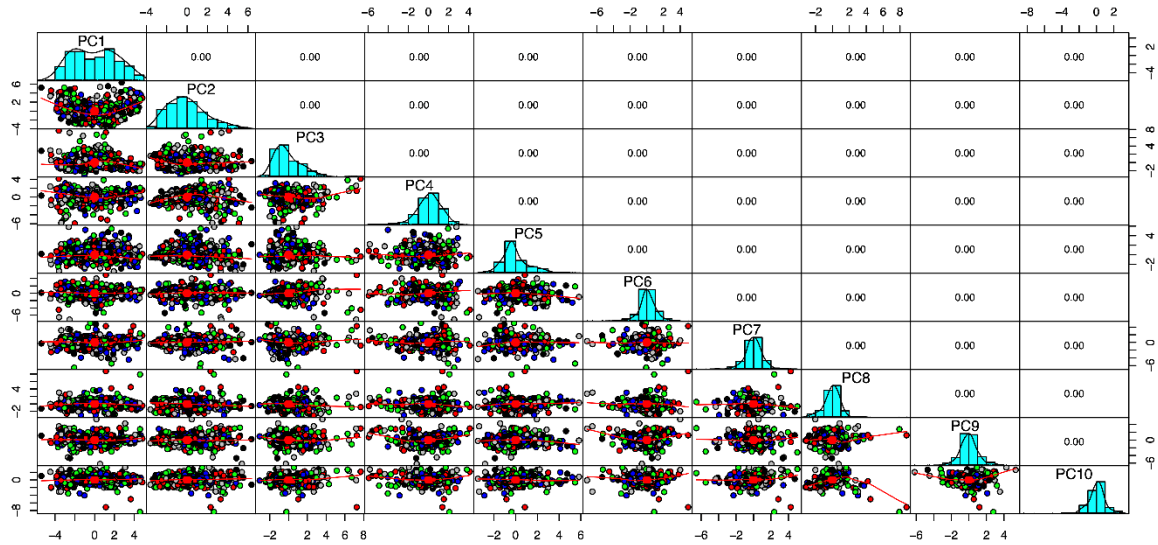
Figures



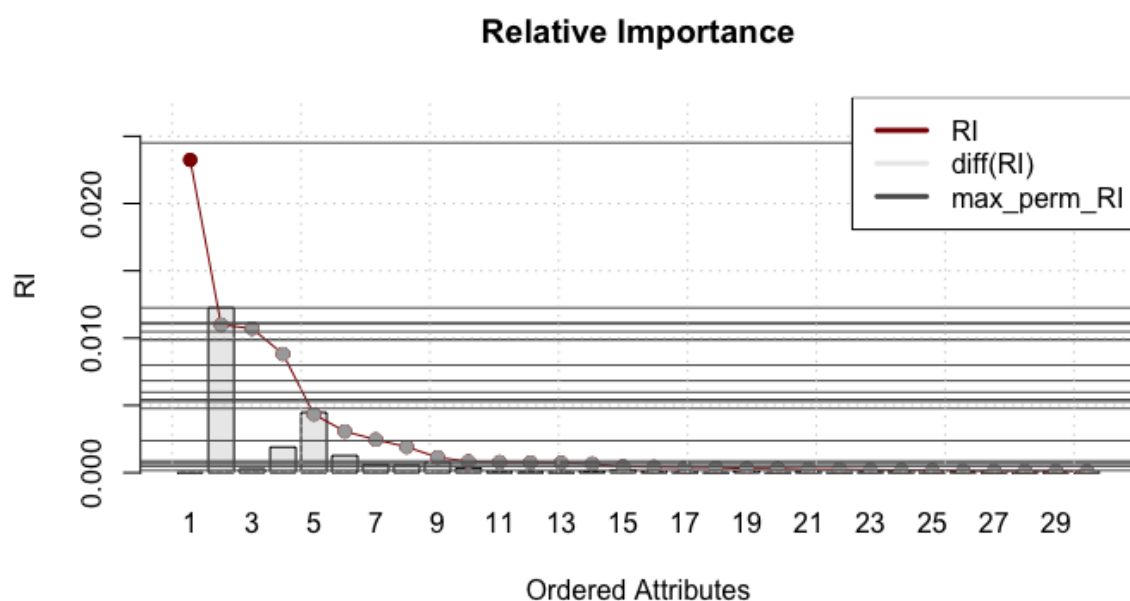
Supplementary Figure 1. ROC curves for PNH algorithm performance. (A) Initial validation based on the cohort of 89 confirmed PNH patients. (B) Final retrospective test performance including the expanded cohort of 141 confirmed PNH patients.



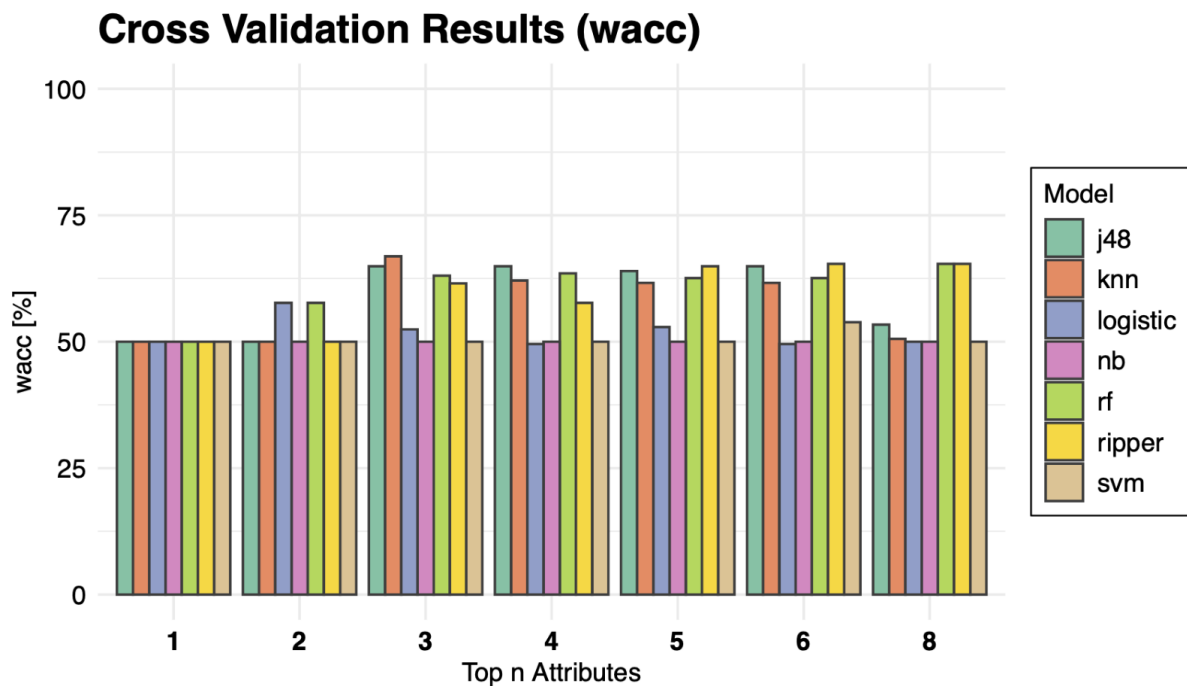
Supplementary Figure 2. Distribution of the odds ratio between paroxysmal nocturnal haemoglobinuria (PNH) positive and negative patients, for being selected by a pair of PNH algorithm variants and obtaining a positive flow cytometry test result. The presented p-value was computed using Fisher's exact test.



Supplementary Figure 3. Relationship between specific principal components and patients assignment to the groups: black-NA: lost to follow up; green-PNH-confirmed: patients whose EHRs were used to train the PNH model; grey-NTD: patients assigned high-risk by the algorithm, but not referred to tests; blue-PNH negative: patients assigned high-risk, who were tested and found negative; red-PNH positive: patients assigned high-risk, who were tested and found positive. Below the diagonal: scatter plots showing the relationship between each pair of variables. On the diagonal: histograms displaying the distribution of each individual variable. Above the diagonal: Pearson correlation coefficients between each pair of variables.



Supplementary Figure 4. Relative importance values for subsequent features differentiating between tested PNH patients who obtained positive vs negative test results. The grey horizontal lines show the RI of the best feature from each of the 20 permutations. The cutoff point based on permutation method returned one significant feature. RI – Relative Importance value; diff (RI) - difference in relative importance between a given and the next feature shown as a grey bar; max_perm_RI – relative importance value of a feature with the highest RI in one of twenty permutation analysis showed as a horizontal grey line.



Supplementary Figure 5. Classification analysis results of tested patients: PNH-positive vs PNH-negative, based on the top features from MCFS-ID ranking, shown as weighted accuracy (wacc) returned in 3-fold cross validation experiment. Classification included patients with positive vs negative paroxysmal nocturnal haemoglobinuria test outcome. Each color represents one classifier: j48 - decision trees; knn - k-nearest neighbors; logistic - logistic regression; nb - Naïve Bayes; rf - random forest; ripper – ripper; svm - support vector machines. Values below the set of bars correspond to the number of top features that were used in the classification experiment.