AI triage for JAK2 testing

Dr Danny Simpson[1], Dr Tara Maisel[2], Dr Shamzah Araf[2], Dr Joseph Taylor[2]

[1]Bristol NHS Foundation Trust
[2]Barts Health NHS Foundation Trust

NHS Fellowship in Clinical AI
NHS Digital Academy Programme

Clinical problem

Myeloid gene panel (MGP) testing
- Range of indications including suspected polycythemia vera (PV)
- 13,000 MGP requests from East London/South East (2019-2025)
- Large backlog

Polycythemia vera workflow
- Referrals based on WHO 2016 criteria for JAK2 testing (via MGP) with ~10% PPV
- Multiple alternative algorithms developed attempting to further rationalize testing

Proposed solution

1 Define dataset

Assess currently available algorithm performance locally

2 Develop binary classifier
- *Features* – FBC, age, gender
- *Target* - JAK2 positivity

Temporal/external validation

3 Implement
- Use predicted probability to prioritise testing
- Request further information for low likelihood referrals

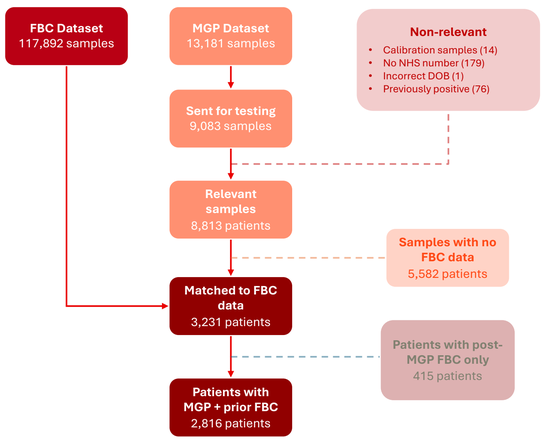
Dataset

MGP and FBC data were paired, with the aim of linking a diagnostic MGP to the closest preceding FBC for all patients

Challenges;
- Unclean data from multiple hospital sites
- Often multiple MGP tests per patient
- Some patients undergo MGP testing whilst on treatment (non-diagnostic samples, FBC influenced by prior treatment)

Data filtered for MGP tests performed in those that met WHO 2016 criteria for JAK2 testing

Final dataset;
718 PV-related samples
88 JAK2 positive



Benchmarking

Four current algorithms designed for prediction of JAK2 positivity in the context of suspected PV identified from the literature
- Reported sensitivities 91-100%, negative predictive values 97.5-100%

Aim to develop algorithm with;
1. Maximum sensitivity – no JAK2 cases missed
2. Clinically useful specificity – able to screen out a relevant number of requests whilst still identifying all JAK2 cases

Chose to assess algorithms using sensitivity and specificity rather than negative predictive value (NPV);
- NPV does not account for false positives and therefore cost of over-calling
- Low prevalence setting of PV means false negatives are relatively rare, and NPV unable to discriminate well between algorithms

No algorithm combined both high sensitivity and specificity in our data
- Only algorithm that achieved a sensitivity of 100% demonstrated near-zero specificity, theoretically able to screen out only 3% of testing requests

	JAK-2 tree[1]	Piris-Villaespesa[2]	Maeyama[3]	JAKPOT[4]
Sensitivity	100%	96%	27%	63%
Specificity	3%	41%	99.5%	67%

Development

1. Multiple model candidates trained on data
- Data split into training and validation cohorts, using a stratified 80:20 train/test split
- Three binary classification models (LogisticRegression, XGBoost and RandomForest classifiers) trained to predict JAK2 status

2. Hyperparameter tuning performed to optimise sensitivity for the JAK2-positive class
- Randomised search approach across a range of hyperparameters
- Repeated stratified cross-fold validation (5 folds, 3 repeats)

3. Decision threshold adjustment
- Hyperparameters fixed, and training data used to generate out-of-fold predictions across 5 cross-folds
- New decision thresholds calculated based on these predictions, selecting thresholds that ensured no false negative predictions throughout all 5 folds

Results

	Logistic regression	XGBoost	RandomForest
Before threshold adjustment			
Sensitivity	89%	78%	83%
Specificity	80%	98%	80%
After threshold adjustment			
Sensitivity	100%	100%	100%
Specificity	53%	31%	15%

- Logistic regression classifier was found to be best performing overall, including after threshold adjustment
- Hypothesised to be due to a relatively small training dataset raising the risk of overfitting by more complex models, with decision boundary collapse then seen in these cases once decision thresholds were altered.

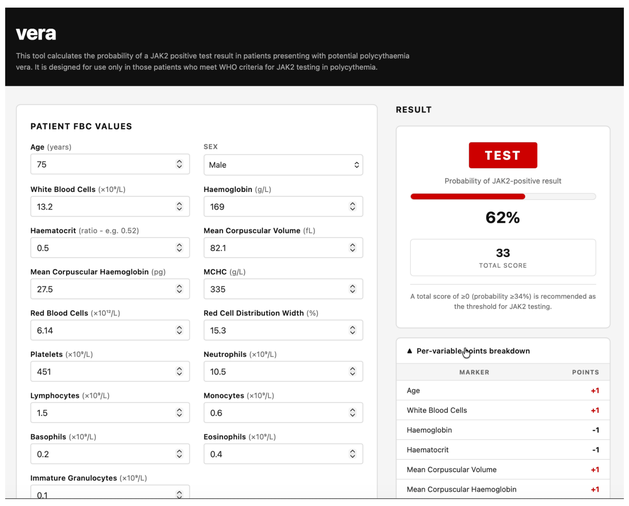
Temporal validation
The best performing model was trained/validated using data up to the end of 2025. Temporal validation using data from the first 6 months of 2026 is now underway – initial results demonstrate sensitivity of 100% is maintained, with specificity of 57%.

External validation
Data from 2023-2026 has been requested from Barking, Havering And Redbridge University Hospitals NHS Trust, which will comprise approximately 400 patients for further validation of model performance in alternative settings.

Implementation

Adaptations to model
- Preparing for implementation by laboratory staff
- Aim to make the model more tangible and easier to demonstrate prior to embedding in workflow

Scorecard-based alternative
- Logistic regressor re-trained using weight-of-evidence approach
- Separate each feature into multiple bins that adds points to overall likelihood score
- Demonstrates how each feature contributes to final prediction in direction and magnitude



Challenges and future work

Challenges
- Ensuring clean dataset with no patients previously diagnosed or treated
- Linking data between sites
- Defining modelling task and expectations (100% sensitivity difficult to guarantee)
- Selecting correct metrics to train for

Future work
- Complete temporal and external validation
- Present adapted model to potential end-users
- Embed within clinical workflow

References
1. Mahe E, Pedersen KM, Çolak Y, Bojesen SE, Lynch T, Sinclair G, Khan F, Shabani-Rad MT. JAK2-tree: a simple CBC-based decision rule to guide appropriate JAK2 V617F mutation testing. J Clin Pathol. 2019 Feb;72(2):172-176. doi: 10.1136/jclinpath-2018-205527
2. Piris-Villaespesa, M., Álvarez-Larrán, A., Saez-Marín, A. *et al.* Development and validation of a sequential two-step algorithm for the screening of individuals with potential polycythaemia vera. *Sci Rep* **11**, 209 (2021). https://doi.org/10.1038/s41598-020-80459-y
3. Keiko Maeyama, Yuka Sugimoto, Keiki Nagaharu, Eiko Oya, Eri Miyata, Atsushi Fujieda, Hiroyuki Miyashita, Takao Sekine, Shigehisa Tamaki, Kohshi Ohishi, Isao Tawara; A Simple Diagnostic Algorithm for the Prediction of *JAK2* Mutation Status in Erythrocytosis Based on Complete Blood Count Data: A Multi-Center Retrospective Analysis. *Blood* 2024; 144 (Supplement 1): 1789. doi: https://doi.org/10.1182/blood-2024-199350
4. Chin-Yee B, Bhai P, Cheong I, Matyashin M, Hsia CC, Kawata E, Ho JM, Levy MA, Stuart A, Lin H, Chin-Yee I, Kadour M, Sadikovic B, Lazo-Langner A. A Rational Approach to JAK2 Mutation Testing in Patients with Elevated Hemoglobin: Results from the JAK2 Prediction Cohort (JAKPOT) Study. J Gen Intern Med. 2023 Jun;38(8):1828-1833. doi: 10.1007/s11606-022-07963-x. Epub 2022 Nov 30. PMID: 36451015; PMCID: PMC10271984.