

Highlights

A Machine Learning Model Built with Time-Dependent Lab Data to Predict Chronic Urticaria Duration and Provide Clinical Insights

Yijun Yang, Guofang Li, Zhen Zhang, Zhirong Yao, Linting Huang, Hui Zhang

- A machine learning model was built to predict the visit duration of urticaria based on clinical laboratory data.
- Inversed trend of shapley values of laboratory data in different phase of disease was observed.

A Machine Learning Model Built with Time-Dependent Lab Data to Predict Chronic Urticaria Duration and Provide Clinical Insights

Yijun Yang^{a,1}, Guofang Li^{a,1}, Zhen Zhang^a, Zhirong Yao^a, Linting Huang^{b,**}, Hui Zhang^{a,*}

^a*Shanghai XinHua Hospital affiliated to Shanghai Jiao Tong University School of Medicine, Address line of institution one, Shanghai, 200092, Shanghai, China*

^b*The ninth hospital of Shanghai Jiao Tong University School of Medicine, Address line of institution two, Shanghai, 200011, Shanghai, China*

Abstract

Background: Chronic spontaneous urticaria (CSU) is a persistent condition characterized by wheals and/or angioedema that significantly impacts patients' quality of life. Predicting the duration of CSU is crucial for effective management and treatment planning. While several biomarkers have been linked to CSU severity and duration, there is limited research on integrating clinical and laboratory data to build predictive models for disease prognosis.

Objective: This study aims to develop a machine learning model to predict visit duration in patients with urticaria based on clinical and laboratory data, and to identify key factors contributing to prolonged episodes of chronic urticaria.

Methods: A retrospective cohort of patients with urticaria from a dermatology clinic was analyzed. Time-independent and time-dependent laboratory data were extracted from electronic medical records. Five machine learning models (XGBoost, Random Forest, Gradient Boosting Machine, AdaBoost, and Support Vector Machine) were trained to predict visit duration, with binary classification thresholds set at 100 and 365 days. Model performance was evaluated using ROC-AUC, accuracy, precision, recall, and F1 score. SHAP analysis was conducted to interpret the contribution of individual features to model predictions, and the Boruta algorithm was used for feature selection.

Results: The XGBoost model with time-dependent data outperformed other models, achieving an AUC of 0.932 for the 100-day threshold. SHAP analysis identified basophils, eosinophils, C-reactive protein (CRP), and immunoglobulin E (IgE) as key predictors of prolonged visit duration, with phase-specific trends in their predictive power. Age-dependent effects were also observed, with certain laboratory markers showing differential impacts in younger versus older patients.

Conclusion: Our machine learning model successfully predicts visit duration in urticaria patients, providing insights into key clinical and laboratory markers that influence disease prognosis. The findings offer a valuable reference for personalized treatment strategies and highlight the importance of phase- and age-specific approaches in managing chronic urticaria.

Keywords:

1. Introduction

Urticaria is a common condition presenting with wheals, angioedema, or both, driven by mast cell degranulation (Zuberbier et al., 2021; Radonjic-Hoesli et al., 2018; Ring and Grosber, 2012). The lifetime prevalence for acute urticaria is approximately 20% (Zuberbier et al., 2021). Urticaria is classified based on duration and triggers. Acute urticaria lasts less than 6 weeks, often triggered by specific causes like drugs, food, or infections. While chronic urticaria lasts more than 6 weeks and can be further classified into chronic spontaneous urticaria (CSU) and chronic

*Corresponding author. Email: authora@example.com

**Corresponding author. Email: huanglinting@163.com

¹Yijun Yang and Guofang Li contributed equally to this work.

inducible urticaria (CIndU)(Zuberbier et al., 2021; Ring and Grosber, 2012). CSU is characterized by the spontaneous occurrence of wheals and/or angioedema without a specific trigger and often associated with autoimmune mechanisms(Schettini et al., 2023), while CIndU is triggered by specific stimuli like cold, heat, or pressure(Pozderac et al., 2020). The prevalence of CSU is approximately 0.5% in general population, and is less prevalent in children compared to adults(Balp et al., 2015; Poddighe, 2019; Labbene and Tekou, 2023).

Chronic urticaria is considered a severe allergic disease with high disease burden(Zuberbier et al., 2021). Epidemiologic studies have established that half of the patients with chronic urticaria achieved remission after 11 months, and approximately 90% achieved remission within 5 years(Eun et al., 2019). Several studies have indicated that the duration and disease severity of CSU are associated. It is observed that longer disease duration and positive ASST were present in patients with more severe symptoms(SanchezBorges et al., 2017). Predicting the duration of chronic urticaria, is crucial for effective patient management and treatment planning.

Several clinical features have been associated with the severity and duration of chronic spontaneous urticaria (CSU). Higher age at onset, female gender, longer disease duration, and hypersensitivity to aspirin or nonsteroidal anti-inflammatory drugs (NSAIDs) are linked to more severe CSU and prolonged time to remission (SanchezBorges et al., 2017; Rabelo-Filardi et al., 2013). Patients exhibiting concomitant CIndU and recurrent angioedema also tend to experience longer durations of CSU (SanchezBorges et al., 2017; Curto-Barredo et al., 2018a). Moreover, patients with multiple allergic conditions are more likely to have prolonged episodes of urticaria (Lin et al., 2011).

Potential biomarkers for CSU severity and duration have been identified. Positive autologous serum skin test (ASST) results, basophil counts, levels of inflammatory markers, activation markers of the extrinsic coagulation pathway, immunoglobulin E (IgE), eosinophils counts, and vitamin D levels are all associated with the disease's severity and duration (SanchezBorges et al., 2017; Rabelo-Filardi et al., 2013; Kolkhir et al., 2019). Specifically, plasma levels of prothrombin fragment, D-dimer, and C-reactive protein (CRP) may serve as markers of CSU severity (Rabelo-Filardi et al., 2013). Serum diamine oxidase (DAO) levels have been linked to the response to antihistamines and dietary interventions, indicating a potential role in predicting disease duration (Chiang et al., 2022).

Metabolic factors also play a role, with high waist circumference (WC), rather than high body mass index (BMI), emerging as a predictive risk factor for longer disease duration in CSU patients (Kim et al., 2021).

The aim of this study was to build a machine learning model for predicting the visit duration of urticaria based on clinical laboratory data and to identify the factors can contribute to prolonged episodes of chronic urticaria by analyzing the importance of variables in the model, hoping to provide a reference for the clinical management of urticaria.

2. Methods

2.1. Patients

Patients with urticaria were recruited from the urticaria specialty clinic of the dermatology department of Shanghai XinHua Hospital affiliated to Shanghai Jiao Tong University School of Medicine from January 2019 to December 2023. The inclusion criteria were as follows: (1) patients diagnosed with urticaria according to the EAACI/GA2LEN/EDF/WAO guidelines(Zuberbier et al., 2021); (2) patients with complete clinical and laboratory data; (3) patients with stable follow-up history, indicated by at least 3 times of follow-up visits. the exclusion criteria were as follows: (1) patients with other skin diseases; (2) patients with severe systemic diseases; (3) patients with incomplete clinical data. The study was approved by the ethics committee of Shanghai XinHua Hospital affiliated to Shanghai Jiao Tong University School of Medicine, and all patients provided written informed consent.

2.2. Data collection and management

Data of patients with urticaria were collected from the electronic medical record system of the hospital, including demographic data, clinical data, laboratory data. The data were stored in a Mysql database for subsequent analysis, with schema as follows: (1) Patients: containing basic information of each unique patient; (2) OutpatientNumbers: storing relationship between outpatient numbers with unique patient; (3) PatientVisits: containing visit events records; (4) PatientExaminations: containing the examination events records; (5) ExaminationItems: a dictionary table describing the examination items. the database schema is shown in database. Database markup language (DBML) describing the database schema is shown in supplementary materials Appendix A

2.3. Feature extraction and feature engineering

The following features were extracted from the database: (1) demographic data: gender, first visit age; (2) clinical data: whether the patient having concomitant inducible urticaria; (3) laboratory data: results from common blood tests, CRP, immunoglobulin, 25-hydroxyvitamin D, Thyroid function, autoantibodies, coagulation function, common urine tests, and allergen specific IgE tests.

For laboratory data, 2 types of data were extracted: time-independent data and time-dependent data. Time-independent data are average values of laboratory data during the whole follow-up period, while time-dependent data are average values of laboratory data during preclinical phase (before the onset of urticaria), acute phase (within 6 weeks after the onset of urticaria), and chronic phase (after 6 weeks of the onset of urticaria).

Two datasets were compared for prediction performance in the model development process. The SQL queries for data are shown in supplementary materials Appendix A and Appendix A. For each dataset, features with serious missing data (missing rate > 95%) were excluded. Patients with over 50% null values with the left features were included in the dataset, and the missing values were imputed with the median value of the feature.

Visiting duration, calculated by the difference between the first visit date and the last visit date in the urticaria specialty clinic, was used as surrogate for remission time of urticaria, the validity of which will be analyzed in results section1. Two binary classification thresholds were set for visiting duration: 100 days and 365 days, based on natural history of chronic urticaria shown in previous studies(Eun et al., 2019; Curto-Barredo et al., 2018b; Gaig et al., 2004).

2.4. Model development and comparison

Dataset was split into training set and external test set with a ratio of 7:3. 5 models were adopted for comparison: Xgboost, random forest, gradient boosting machine (GBM), adaboost, and support vector machine (SVM). Time-dependent data extraction may introduce multicollinearity between features, which can affect model performance. To address this issue, L1 regularization, L2 regularization, and column sampling methods were employed in XGBoost to handle multicollinearity. L1 regularization helps eliminate irrelevant collinear features by sparse feature selection, while L2 regularization reduces the model's dependence on collinear features by smoothing the weights. Column sampling reduces the likelihood of collinear features being selected simultaneously by randomly choosing a subset of features when building each tree, which is helpful when interpreting the model using method like SHAP, since it enables more balanced weight distribution across features and reduces the possibility of one feature overshadowing others. The Random Forest model does not support L1 and L2 regularization, but it handles multicollinearity through the max_features parameter, which is similar to column sampling. L2 regularization was applied to the SVM model as well. Additionally, the frameworks of XGBoost and GBM incorporate greedy split and boosting mechanisms, selecting the most important features for each tree and fitting the residuals step by step, allowing the model to focus on the key features rather than over-relying on collinear ones.

Hyperparameter optimization on 10 model settings (5 models and 2 datasets) was performed using TPE algorithm by nni package in python, which is a bayesian optimization algorithm that uses tree-structured parzen estimator to model the objective function and suggest the next set of hyperparameters to evaluate based on the previous results. Internal 5 fold cross-validation was employed to discern the most suitable hyperparameters for each distinct model, individually applied to each model settings for enhanced performance. The model's performance was assessed using ROC-AUC, accuracy, precision, recall and F1 score on different cutoffs of visiting duration. The sensitivity, specificity, PPV, NPV, accuracy, and F1 score were calculated at the optimal cutoff value that maximized the Youden index. The best-performing model and data were selected for further analysis.

2.5. feature selection

Too many features can lead to overfitting and reduce the interpretability of the model. Therefore, feature selection was performed on final model for further optimization. Although, feature importances can be evaluated directly from the boosted trees, these importances have been shown to be local and inconsistent. A SHAP score inspired by Shapley values can combine different explanation models and provide a global feature importance score that is consistent across different test sets. However, other than to arbitrarily select an importance threshold beyond which features are considered unimportant, SHAP analysis does not offer an algorithmic way to filter a large feature set to a limited set of important features. Boruta algorithm is a wrapper method that determine the importance of features by comparing the

importance of original features with their shuffled copies. If importance of an original feature is significantly greater than its shuffled copy, that feature is deemed important (Kursa and Rudnicki, 2010).

In this study, Boruta algorithm was performed on the final model, with max iteration set to 50 and repetition for 15 times. The confirmed features list and importance rank were recorded for each run. The overlap of confirmed features across runs was calculated to determine the final feature list. For time-dependent data, once a laboratory item was confirmed as important in any phase of disease, the feature of the same lab item in other phase of disease was also considered as important. That is to say, when top 10 items were selected, the features representing different phases of that item were all included in the final model, which is at most 30 features for time-dependent data. The model performance on different number of top items, ranked by the average importance score, were compared to determine the optimal number of features for the final model.

2.6. Model explanation

SHAP (Shapley Additive Explanations) is an explanation method based on game theory, used to interpret the outputs of machine learning models. It assigns each feature's contribution to the model's prediction, helping users understand the decision-making process of the model. In tree models that are not sensitive to collinearity, such as Xgboost and GBM, the model tends to select one most important feature among collinear features for a split and ignores the other correlated features. As a result, SHAP may assign a high contribution to the first selected feature while the others receive little or no contribution. This happens because the model captures most of the information through the first feature, effectively overshadowing the others. By introducing joint distribution estimation for correlated features during SHAP calculations, we can capture their dependencies and ensure that SHAP values reflect the true impact of correlated features (Lundberg and Lee, 2017). The shapley value of each laboratory item in different phase of disease were compared to reveal various predicting ability. Some of the features were further analyzed with different age populations to explore the potential age effect on the predicting ability of laboratory data. Kernel density estimation was used to visualize the distribution of laboratory data in different phase of disease to provide a reference for the interpretation of SHAP values.

3. Results

3.1. Visit duration as a validate surrogate for time to remission

To verify visit duration as a surrogate for natural time to remission, we analyzed the distribution of visit duration in the dataset. The distribution of visit duration is shown in Figure 1. Similar to epidemiologic study which reported that half of the patients with chronic urticaria achieved remission after 11 months, and approximately 90% achieved remission within 5 years (Eun et al., 2019), the distribution of visit duration in our dataset showed that approximately 68% of all patients have a visit duration more than 6w. Within those patients, 54% of them have a visit duration less than a year, and rest of them have a visit duration more than a year. Unfortunately, our dataset does not include any patients with a visit duration exceeding five years. This study's visit duration distribution reflects the chronic urticaria remission pattern, suggesting that visit duration serves as a reliable proxy for remission time.

3.2. Patient characteristics

Of 9921 patients recruited for this study, 3954 patients were selected for time-independent dataset and 1948 patients were selected for time-dependent dataset, the rest were excluded due to serious missing data as described in methods section. Two datasets were allocated into separate derivation and external validation sets with a ratio of 7:3, respectively. The comparison of demographic and clinical variables among the training and external validation sets is shown in Table Appendix A for time-independent dataset and Table Appendix A for time-dependent dataset. No substantial differences were observed between the training set and the external test set across the majority of features in either time-independent or time-dependent data. Details of the study design are displayed in Figure 2.

The comparison of clinical characteristics between patients with short and prolonged visit duration in the time-independent dataset (Table 1) highlights several trends that align with findings from previous studies on chronic spontaneous urticaria (CSU). Patients with visiting duration more than 100 days were defined as poor outcome group, and patients with visiting duration less than 100 days were defined as good outcome group. In this analysis, immunoglobulin E (IgE) (131.65 ± 242.00 in good outcome vs 152.28 ± 346.75 in poor outcome), and basophil percentage (0.25 ± 0.22

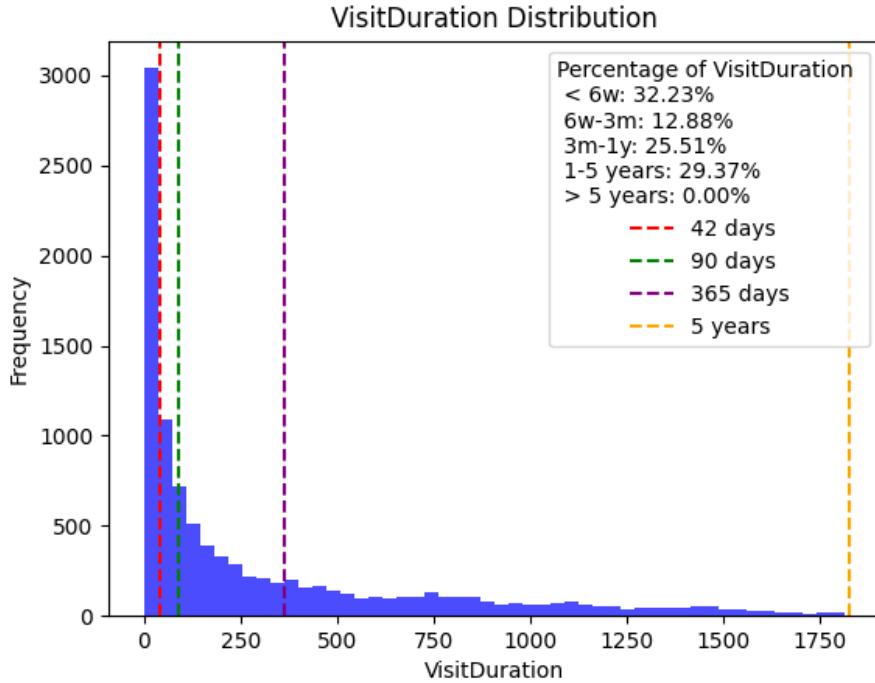


Figure 1: Distribution of visit duration in the dataset. The x-axis represents the visit duration in days, and the y-axis represents the number of patients. Lines with different colors represent different cutoffs of visit duration. The percentage of patients within different cutoffs are shown in the legend.

in good outcome vs 0.32 ± 0.24 in poor outcome) were significantly higher in patients with poor outcomes, which is consistent with previous studies that have linked these markers to CSU severity and duration (SanchezBorges et al., 2017; Rabelo-Filardi et al., 2013).

In contrast, C-reactive protein (CRP) levels were significantly lower in patients with poor outcomes (13.03 ± 19.62 in good outcome vs 7.71 ± 10.23 in poor outcome), which contradicts previous findings that have associated higher CRP levels with CSU severity (Rabelo-Filardi et al., 2013). However, this discrepancy could reflect the complexity of CRP as an inflammatory marker that may vary across different disease stages or subgroups of patients. In time-dependent data, which is shown in supplementary Table Appendix A, CRP presented various trends in different phases of disease, with significantly lower CRP levels in acute phase (10.21 ± 10.95 in good outcome vs 8.95 ± 8.22 in poor outcome) and higher CRP levels in chronic phase (5.39 ± 4.84 in good outcome vs 6.47 ± 8.81 in poor outcome) in patients with poor outcomes. This finding suggests that the predictive ability of CRP may be influenced by the disease phase, which could explain the inconsistent results observed in the time-independent dataset.

3.3. Comparison of multiple models on time-dependent and time-independent data

Time-independent and Time-dependent data, combined with 5 different models (XGBoost, Random Forest, AdaBoost, GBM, SVM), were used to generate 10 machine learning models to predict the visiting duration of urticaria. Among the 10 models, the XGBoost model with time-dependent data ($AUC = 0.932 \pm 0.016$ for 100, $AUC = 0.860 \pm 0.028$ for 365) has the best predictive effect for visiting duration, followed by the Random Forest model with time-dependent data ($AUC = 0.922 \pm 0.028$ for 100, $AUC = 0.837 \pm 0.037$ for 365) as shown in Figure 3.

Notably, time-dependent data consistently outperformed time-independent data across all models, indicating that the phase-specific laboratory data are more informative for predicting the visiting duration of urticaria.

The discriminative performances and ROC curves of all 10 models are listed in Supplementary Table Appendix A, and Supplementary Figure Appendix A, respectively.

Characteristic	Good Outcome	Poor Outcome	P-value
Number of patients	2027	1927	
Outcome	25.37 ± 25.09	604.32 ± 437.04	0.000 ↑
Gender	0: 64.7%, 1: 35.3%	0: 61.1%, 1: 38.9%	0.023
First Visit Age	30.12 ± 21.79	28.11 ± 23.61	0.005 ↓
CI nd U	0: 99.3%, 1: 0.7%	0: 97.2%, 1: 2.8%	0.000
Lymphocytes Percentage	28.98 ± 14.05	34.02 ± 13.88	0.000 ↑
Neutrophils Percentage	63.13 ± 15.68	56.63 ± 14.97	0.000 ↓
Monocytes Percentage	6.18 ± 1.64	6.49 ± 1.67	0.000 ↑
Mean Corpuscular Hemoglobin Concentration	335.62 ± 10.51	336.93 ± 10.30	0.000 ↑
Platelet Count	263.47 ± 74.09	256.73 ± 68.96	0.003 ↓
White Blood Cell Count	10.31 ± 3.00	9.35 ± 2.39	0.000 ↓
Mean Corpuscular Hemoglobin	29.07 ± 2.33	28.98 ± 2.30	0.202
Mean Corpuscular Volume	86.66 ± 6.65	85.99 ± 6.67	0.002 ↓
Hemoglobin	130.03 ± 9.16	130.20 ± 8.46	0.546
Eosinophils Percentage	1.55 ± 2.02	2.31 ± 2.37	0.000 ↑
Basophils Percentage	0.25 ± 0.22	0.32 ± 0.24	0.000 ↑
Absolute Eosinophil Count	0.13 ± 0.20	0.18 ± 0.21	0.000 ↑
Absolute Lymphocyte Count	2.61 ± 1.12	2.83 ± 1.29	0.000 ↑
Mean Platelet Volume	9.14 ± 1.39	9.02 ± 1.39	0.007 ↓
Platelet Distribution Width	12.72 ± 1.96	12.57 ± 1.94	0.020
Eosinophil Count Absolute	128.60 ± 141.68	162.83 ± 190.02	0.000 ↑
C-reactive Protein	13.03 ± 19.62	7.71 ± 10.23	0.000 ↓
Immunoglobulin E	131.65 ± 242.00	152.28 ± 346.75	0.029
SMRNP	1.18 ± 2.16	1.22 ± 1.84	0.576
Anti SSA	1.42 ± 5.44	1.58 ± 5.68	0.363
Anti Jo 1	1.08 ± 1.77	1.12 ± 2.06	0.534
Nucleosome	0.57 ± 0.34	0.65 ± 0.43	0.000 ↑
Ribosomal PP rotein	1.07 ± 0.71	1.17 ± 2.13	0.050
Ro 52	2.10 ± 5.90	2.23 ± 6.81	0.515

Table 1: Comparison of the characteristics between patients with good and poor outcomes in the time-independent dataset
 continuous variables are presented as mean ± standard deviation, categorical variables are presented as number (percentage)
 good outcome is defined as visit duration < 100 days, poor outcome is defined as visit duration ≥ 100 days

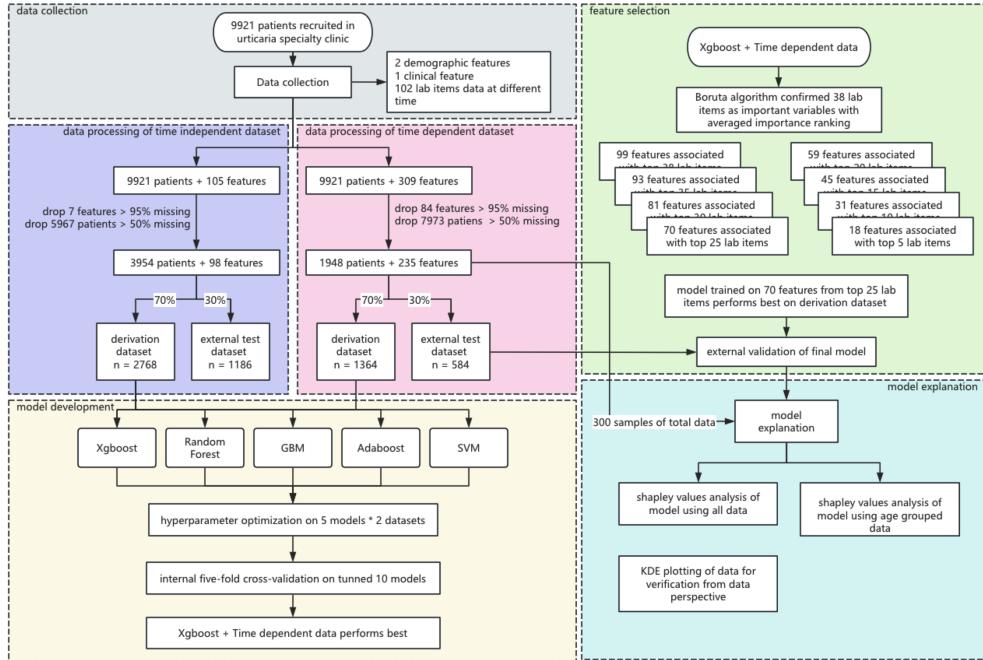


Figure 2: Study design

3.4. Feature selection and final model

Xgboost with time-dependent data, as the best performing model, was selected for further optimization. The Boruta algorithm identified 38 laboratory items as important, as shown in Supplementary Table Appendix A. Feature importance ranking by Boruta algorithm are shown in Figure 4 and Supplementary Table Appendix A. We rank the confirmed features by the average importance score and compare the model performance on different number of top items.

During the process of reducing features based on the feature importance, the changes in AUCs for the model shows that the model maintain a robust performance until the number of items reduced to less than 25, as shown in Figure 5 A.

The 25-item model were selected as the final model, and showed a stable performance in external test set with AUC of 0.956 and 0.877 for 100 and 365 cutoffs, as shown in Figure 3 5 B.

3.5. Model explanation

SHAP analysis was conducted on the final model to evaluate the feature importance. Figure 6 presents a heatmap of SHAP values from 100 randomly selected samples, ordered by hierarchical clustering to group similar patterns. The heatmap illustrates how these features influence the model's output, with global feature importance displayed as a bar plot.

3.5.1. SHAP analysis reveals alignment with previous studies while discovering phase-specific trends

SHAP analysis on well-known markers of CSU severity and duration, such as eosinophils, basophils, and CRP, revealed consistent trends with previous studies while providing insights on how those markers in different phase can have different predicting ability.

A high basophil percentage during the chronic phase of the disease pushed the model output towards predicting a longer visit duration. This finding aligns with previous research on chronic spontaneous urticaria (CSU) (SanchezBorges et al., 2017). Interestingly, basophils had no impact in the acute phase and showed a reversed effect during the preclinical phase (Figure 7 A).

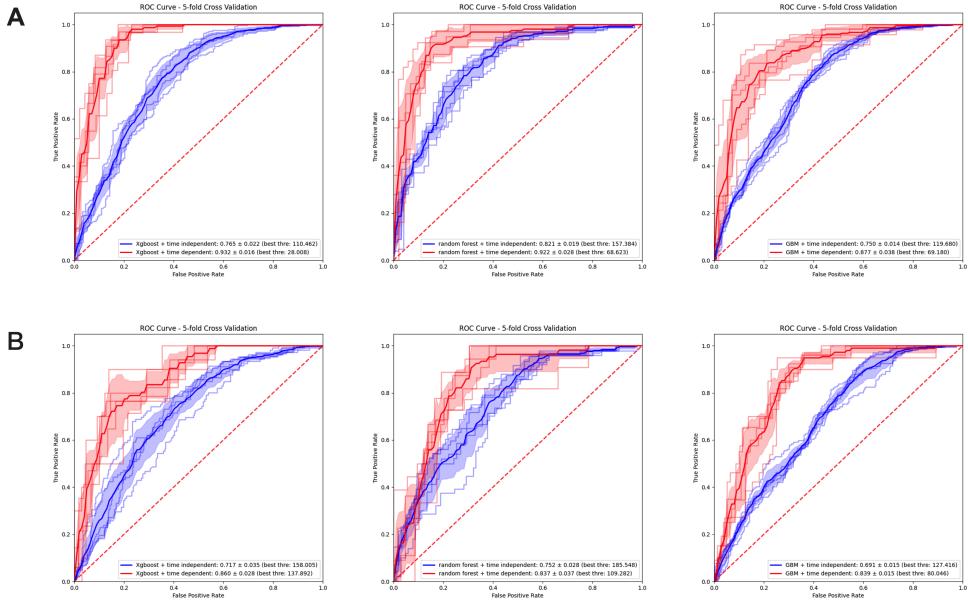


Figure 3: ROC curves for 5-fold cross-validation results for three different models (XGBoost, Random Forest, and GBM) at 2 binary classification thresholds (100 and 365) in panel A and B. The models' performances are plotted with time-independent and time-dependent datasets, where red indicates the time-independent dataset and blue represents the time-dependent dataset. The shaded areas represent the 95% confidence intervals.

A low eosinophil percentage or count consistently pushed the model toward predicting longer visit durations across all disease phases (Figure 7 B). This finding is in line with previous studies, which associate eosinopenia with type IIb autoimmunity, heightened disease activity, and poor treatment responses in CSU patients (Kolkhir et al., 2019).

Lymphocyte percentage and neutrophil percentage in chronic phase also shows harmful effect on patients with longer visit duration, as shown in Figure 7 C and D.

High immunoglobulin E levels in the chronic phase of disease was having positive effect on the model output, which is consistent with previous studies that have linked high IgE levels to CSU severity and duration (SanchezBorges et al., 2017), while having no effect in the acute phase of disease, as shown in Figure 7 C.

The increasing CRP levels in the chronic phase of disease was having positive effect on the model output, same as the previous studies that have associated higher CRP levels with CSU severity (Rabelo-Filardi et al., 2013), while the extreme high CRP levels in the acute phase of disease was having negative effect on the model output, as shown in Figure 7 B.

High mean platelet volume was having positive effect on the model output, a trend that was proved by previous studies (Aleem et al., 2015; Magen et al., 2010), while having no effect in the acute and preclinical phase of disease, as shown in Figure 7 G.

Thyroid dysfunction is a common comorbidity in patients with chronic urticaria, and it has been reported that reducing thyroxine dose may cause flare-ups of chronic urticaria and angio-oedema, suggesting a possible link between thyroid function and urticaria severity (Dunkley and Jawad, 2003). Our study shows that a high TSH level in the chronic phase of disease is having negative effect on the model output, as shown in Figure 7 E.

At last, autoimmunity markers, such as Anti Jo 1 and Ro 52, are having positive effect on the model output, as shown in Figure 7 I and J.

Other features that were not mentioned in the previous studies, such as mean corpuscular hemoglobin concentration, mean corpuscular volume, and platelet distribution width, also showed significant effects on the model output, as shown in Supplementary Figure Appendix A.

KDE plots of features mentioned above in different phases of disease are shown in Supplementary Figure Appendix A to illustrate the feature effect from the data perspective.

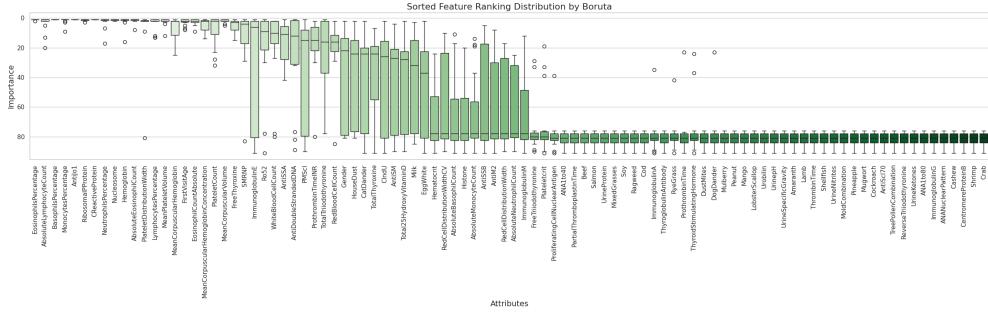


Figure 4: Boxplot representing averaged feature ranking distribution for each laboratory item generated by the Boruta algorithm over 15 iterations. The ranking value for each laboratory item is averaged for 3 different phases of disease. The y-axis shows the importance rank, with lower values indicating higher importance.

3.5.2. SHAP analysis in different age groups shows novel age effect

Age-specific analysis revealed that the predictive power of laboratory data varies across different age groups. Notably, the absolute eosinophil count in the acute phase had a stronger positive effect on the model's output for patients aged 0-2, while this effect diminished in patients older than six years (Figure 8).

Absolute lymphocyte count also exhibited an age-dependent effect. SHAP analysis showed that a high lymphocyte count during the acute and preclinical phases was beneficial, while it was harmful during the chronic phase. Age subgroup analysis revealed that this predictive trend was significant only in patients older than six years, with little to no effect in younger age groups (Figure Appendix A).

4. Discussion

Our study presents several significant findings regarding the prediction of visit duration in patients with chronic spontaneous urticaria (CSU). Utilizing a combination of machine learning models and SHAP analysis, we were able to identify key clinical and laboratory markers that influence disease duration across various phases of CSU. Notably, time-dependent data consistently outperformed time-independent data in predicting visit duration, suggesting that incorporating temporal dynamics provides more nuanced insights into disease progression. Among the models tested, XGBoost with time-dependent data demonstrated the best predictive performance, further highlighting the importance of accounting for longitudinal changes in patient data.

Our comparison of machine learning models demonstrated that the XGBoost algorithm, particularly with time-dependent data, outperformed other models such as Random Forest and GBM. The superior performance of XGBoost can be attributed to its ability to handle non-linear interactions and its robust handling of complex, high-dimensional datasets like those generated in this study. The time-dependent nature of the dataset was particularly important for capturing disease dynamics over time, as patients with CSU often experience fluctuations in their clinical markers and disease activity. This phase-specific approach is crucial, as static time-independent data may overlook these critical variations, resulting in less accurate predictions.

Our findings have significant clinical implications, particularly in the context of personalized medicine. By accurately predicting visit duration based on laboratory and clinical markers, clinicians can better tailor treatment plans, potentially improving patient outcomes. For instance, identifying patients likely to experience prolonged disease duration early in the course of their illness could prompt more aggressive treatment strategies, mitigating disease severity and improving quality of life.

The SHAP analysis provided valuable insights into the importance of specific laboratory markers in predicting visit duration, revealing both phase-specific and age-specific trends. Key markers such as basophils, eosinophils, and CRP were consistently highlighted across different disease phases, aligning with findings from previous studies on CSU severity and duration (SanchezBorges et al., 2017; Kolkhir et al., 2019).

Basophil percentage, for example, was a strong predictor of longer visit duration in the chronic phase but had no effect in the acute phase and a reversed effect in the preclinical phase. This phase-dependent variation suggests that

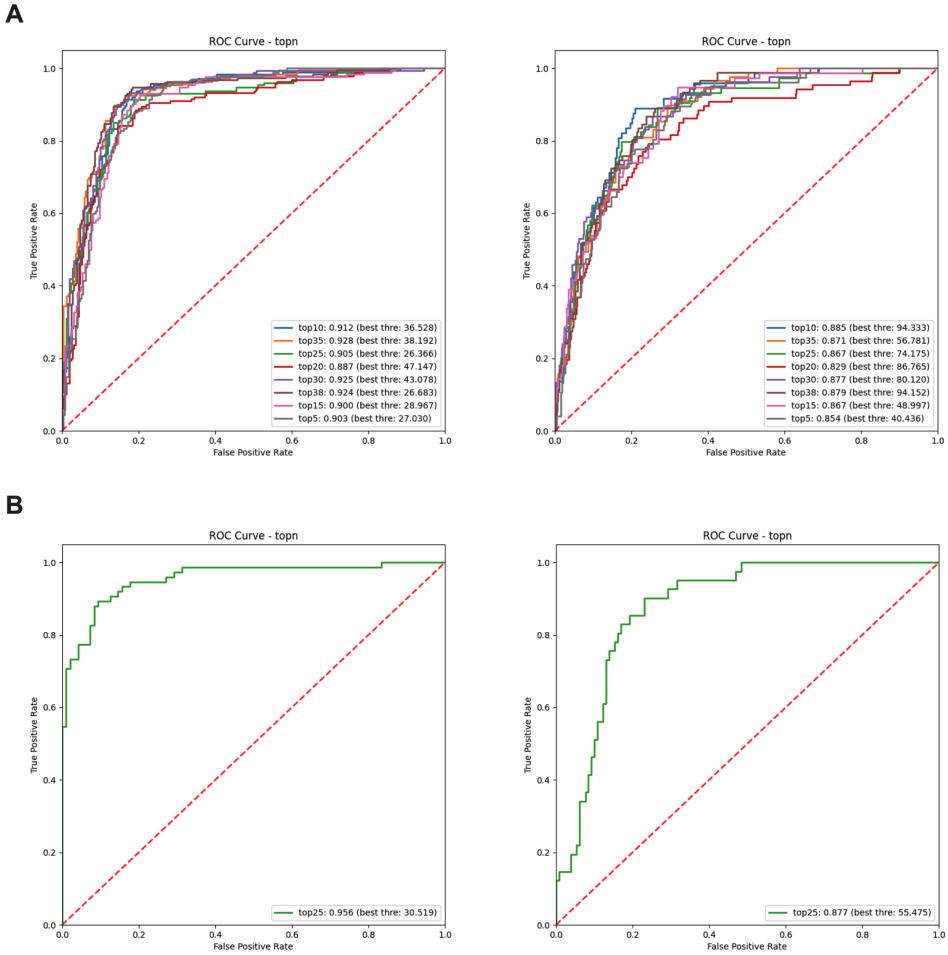


Figure 5: Panels A display the ROC curves for the model as the number of top items is reduced based on the average importance score at cutoff values of 100 and 365, respectively. Panels B show the ROC curves for the final model on external test set at cutoff values of 100 and 365, respectively.

basophils may play different roles at various stages of the disease, with higher basophil levels potentially indicating a more severe and prolonged disease course in chronic CSU. These findings are consistent with prior research linking elevated basophils to disease severity (SanchezBorges et al., 2017).

Eosinophils also emerged as an important predictor, with low eosinophil percentages or counts correlating with longer disease duration across all phases. Eosinopenia has been associated with type IIb autoimmunity, which is characterized by more severe disease and poorer treatment response in CSU patients (Kolkhir et al., 2019). This finding reinforces the clinical utility of eosinophils as a biomarker for CSU prognosis.

CRP levels, another well-established inflammatory marker, exhibited complex, phase-specific effects. In the chronic phase, higher CRP levels predicted longer visit durations, consistent with previous studies linking elevated CRP to greater CSU severity (Rabelo-Filardi et al., 2013). However, in the acute phase, extremely high CRP levels had a negative impact on the model's predictions, suggesting that CRP's role as a biomarker may vary depending on the disease stage. This highlights the importance of interpreting inflammatory markers like CRP in the context of disease progression.

In addition to phase-specific trends, our SHAP analysis revealed novel age-dependent effects in the prediction of visit duration. For example, the predictive ability of absolute eosinophil count in the acute phase was more pronounced

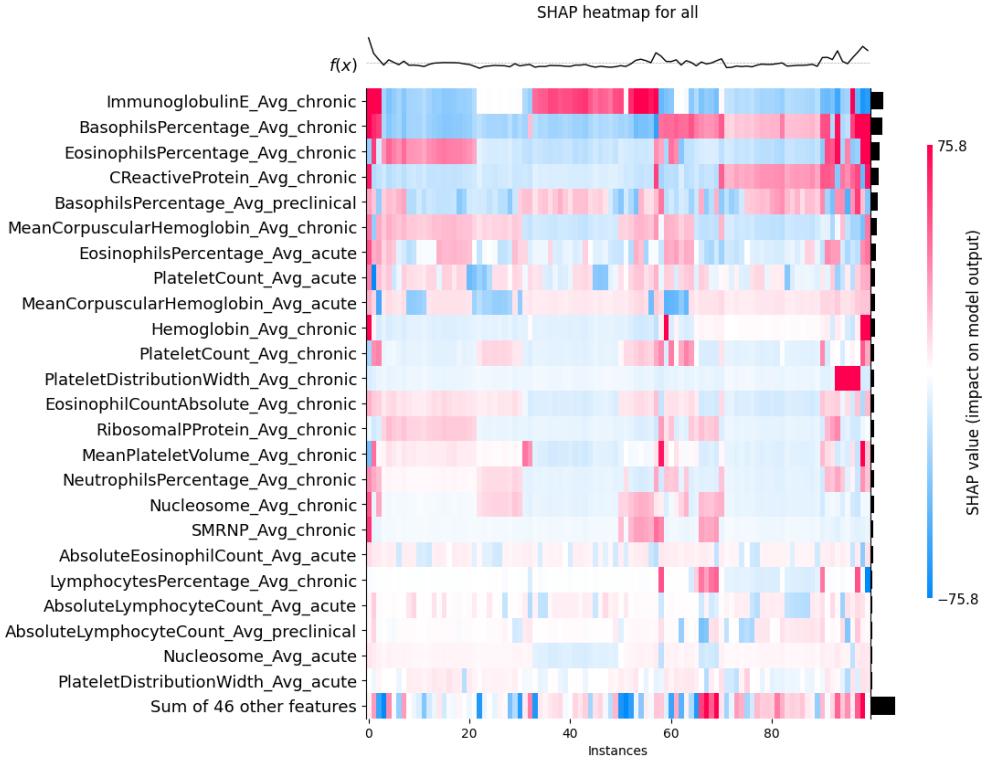


Figure 6: a heatmap plot of SHAP values of all features on 100 randomly selected instance, with the instances on the x-axis, the model inputs on the y-axis, and the SHAP values encoded on a color scale. The samples are ordered based on a hierarchical clustering by their explanation similarity. This results in samples that have the same model output for the same reason getting grouped together. The output of the model is shown above the heatmap matrix, and the global importance of each model input shown as a bar plot on the right hand side of the plot.

in younger patients (aged 0-2) but diminished in those older than six years. Similarly, the impact of absolute lymphocyte count was only significant in patients older than six years, suggesting that the immune response in CSU may differ based on age. These findings underscore the need for age-specific considerations when interpreting laboratory data and highlight the potential for more tailored, pediatric-specific treatment strategies in younger CSU patients.

Our findings build upon and extend previous research on the role of laboratory markers in CSU. Prior studies have identified various biomarkers, such as elevated IgE and eosinophil counts, as indicators of disease severity and duration (SanchezBorges et al., 2017; Kolkhir et al., 2019). However, few studies have explored the phase-specific predictive power of these markers or examined how their relevance may change over the course of the disease. By utilizing SHAP analysis, we were able to uncover these nuanced relationships, demonstrating that certain markers may only be predictive during specific phases of CSU.

Additionally, our study is one of the first to investigate age-dependent effects in CSU biomarker prediction. Previous research has largely focused on adult populations, with little attention given to how biomarkers may function differently in pediatric patients. The age-specific trends we identified offer new avenues for research and suggest that treatment strategies for CSU may need to be adjusted based on patient age.

While our study provides valuable insights into the predictive modeling of visit duration in CSU, several limitations should be acknowledged. First, the lack of patients with visit durations exceeding five years limits the generalizability of our findings to longer-term CSU cases. Future studies should aim to include a broader range of follow-up times to better capture the full spectrum of disease duration. Additionally, while SHAP analysis provides important information about feature importance, it does not establish causality. Further research is needed to explore the mechanistic underpinnings of the relationships we observed between laboratory markers and CSU duration.

Another limitation is the exclusion of patients with missing data, which may have introduced selection bias. Future research should investigate methods for handling missing data, such as multiple imputation, to ensure that results are

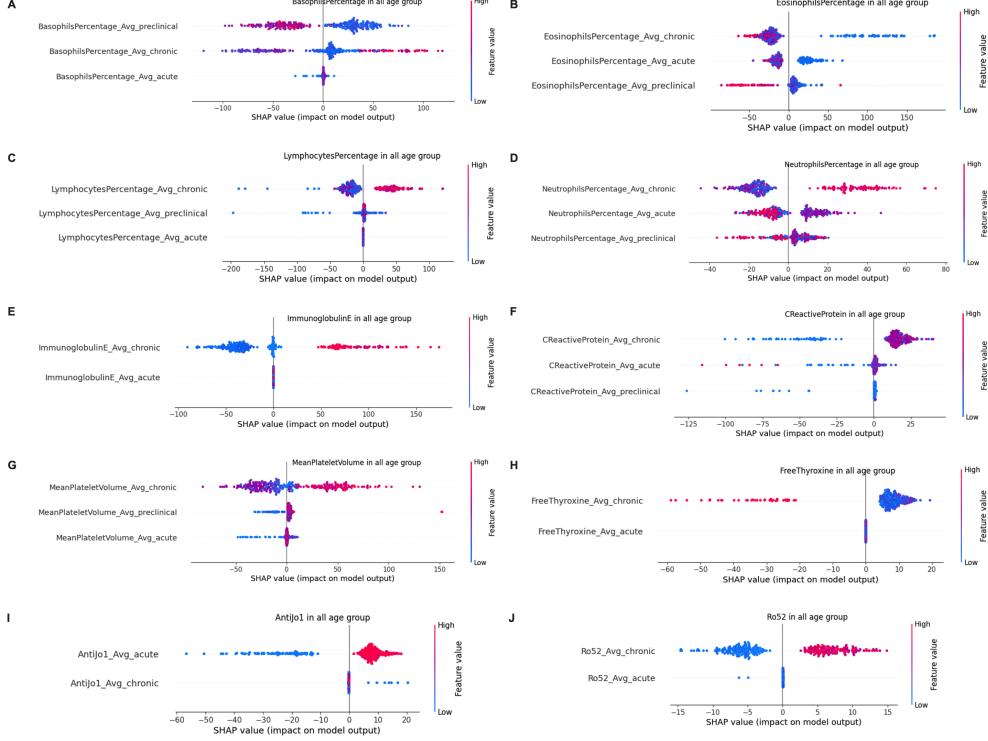


Figure 7: SHAP value summary plots for well-known laboratory markers across different disease phases (preclinical, acute, and chronic) in all age groups. Each dot represents a patient sample, with the x-axis indicating the SHAP value (the feature's impact on the model output) and the color gradient representing the feature value (red for high values, blue for low values). Features with higher SHAP values indicate a greater positive impact on predicting longer visit durations, while negative SHAP values correspond to shorter visit durations.

generalizable to a broader population of CSU patients.

Finally, while our study focused on laboratory markers, future studies should consider incorporating additional clinical data, such as treatment response and comorbidities, to further improve model accuracy and clinical applicability.

In summary, our study highlights the utility of machine learning models, particularly XGBoost, in predicting visit duration in CSU patients. Time-dependent data proved to be critical in capturing the dynamic nature of the disease, while SHAP analysis revealed phase-specific and age-dependent trends in key laboratory markers. These findings not only enhance our understanding of CSU pathophysiology but also have practical implications for personalized treatment strategies. Future research should continue to explore the use of machine learning and longitudinal data in chronic disease management to further optimize patient care.

5. code availability

Code for data extraction, model development, and model explanation are available at <https://github.com/jabberwockyang/UrticariaPrediction>

Appendix A. Supplementary data

Supplementary file 1: Database schema

Supplementary file 2: SQL queries for time-independent extraction

Supplementary file 3: SQL queries for time-dependent extraction

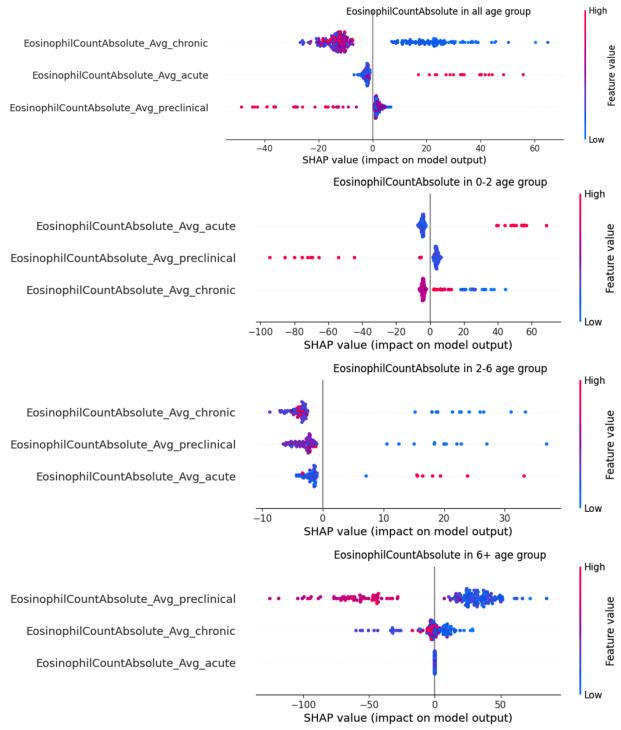


Figure 8: SHAP value summary plots for absolute eosinophil count across different disease phases (preclinical, acute, and chronic) and age groups (all ages, 0-2, 2-6, and 6+ years). Each dot represents a patient sample, with the x-axis indicating the SHAP value (the feature's impact on the model output) and the color gradient representing the feature value (red for high values, blue for low values).

Supplementary Table1: Comparison of the characteristic between the training set and external testing set data in the time-independent dataset

Supplementary Table2: Comparison of the characteristic between the training set and external testing set data in the time-dependent dataset

Supplementary Table 3: Comparison of the characteristic between the short duration group and long duration group in the time-dependent dataset

Supplementary Table 4: The discriminative performances of all 10 models in 5 fold cross validation

Supplementary Figure 1: ROC curves of all 10 models in 5 fold cross validation

Supplementary Table 5: Feature importance ranking by Boruta algorithm

Supplementary Table 6: 38 laboratory items confirmed as important by Boruta algorithm

Supplementary Table 7: The performance of the final model on external test set

Supplementary Figure 2: SHAP analysis on Absolute lymphocytes count in different age groups

Supplementary Figure 3: SHAP value summary plots for other laboratory markers

Supplementary Figure 4: KDE plots of features in different phases of disease

Appendix B. Acknowledgements

References

- Aleem, S., Masood, Q., Hassan, I., 2015. Correlation of mean platelet volume levels with severity of chronic urticaria. *Journal of Dermatology and Dermatologic Surgery* 19, 9–14. doi:10.1016/J.JDDS.2014.03.006.
- Balp, M., Vietri, J., Tian, H., Isherwood, G., 2015. The impact of chronic urticaria from the patient's perspective: A survey in five european countries. *The Patient* 8, 551–558. doi:10.1007/s40271-015-0145-9.
- Chiang, H.L., Chen, C.H., Koo, M., Tsai, T.Y., Wu, C.H., 2022. Predictors of response to oral medications and low-histamine diet in patients with chronic urticaria. *Journal of Immunology Research* 2022. doi:10.1155/2022/5243825.

- Curto-Barredo, L., Archilla, L.R., Vives, G., Pujol, R., Giménez-Arnau, A., 2018a. Clinical features of chronic spontaneous urticaria that predict disease prognosis and refractoriness to standard treatment. *Acta dermatovenereologica* 98 7, 641–647. doi:10.2340/00015555-2941.
- Curto-Barredo, L., Riba Archilla, L., Roura Vives, G., Pujol Vallverdú, R.M., Giménez Arnau, A.M., 2018b. Clinical features of chronic spontaneous urticaria that predict disease prognosis and refractoriness to standard treatment. *Acta Dermato Venereologica*, 641–647.
- Dunkley, L., Jawad, A.S.M., 2003. Thyroid function in chronic urticaria and angio-oedema. *Journal of the Royal Society of Medicine* 96 11, 571. doi:10.1258/JRSM.96.11.571.
- Eun, S.J., Lee, J.Y., Kim, D.Y., Yoon, H.S., 2019. Natural course of new-onset urticaria: results of a 10-year follow-up, nationwide, population-based study. *Allergology International* 68, 52–58.
- Gaig, P., Olona, M., Lejarazu, D.M., Caballero, M., Domínguez, F., Echecipia, S., Abujeta, J.G., Gonzalez, M., Lleonart, R., Cocera, C.M., 2004. Epidemiology of urticaria in spain. *Journal of Investigational Allergology and Clinical Immunology* 14, 214–220.
- Kim, Y., Han, K.D., Bang, C., Lee, J.H., Lee, J.Y., Park, Y., Park, Y., 2021. High waist circumference rather than high body mass index may be a predictive risk factor for the longer disease duration of chronic spontaneous urticaria. *Scientific Reports* 11. doi:10.1038/s41598-021-81484-1.
- Kolkhir, P., Church, M., Altrichter, S., Skov, P., Hawro, T., Frischbutter, S., Metz, M., Maurer, M., 2019. Eosinopenia, in chronic spontaneous urticaria, is associated with high disease activity, autoimmunity and poor response to treatment. *The journal of allergy and clinical immunology. In practice* doi:10.1016/j.jaip.2019.08.025.
- Kursa, M.B., Rudnicki, W.R., 2010. Feature selection with the boruta package. *Journal of statistical software* 36, 1–13.
- Labbene, I., Tekou, A., 2023. Prevalence and treatment of chronic spontaneous urticaria in children. *Scholars Journal of Applied Medical Sciences* doi:10.36347/sjams.2023.v11i09.025.
- Lin, Y.R., Liu, T.H., Wu, T.K., Chang, Y.J., Chou, C., Wu, H.P., 2011. Predictive factors of the duration of a first-attack acute urticaria in children. *The American journal of emergency medicine* 29 8, 883–9. doi:10.1016/j.ajem.2010.04.004.
- Lundberg, S., Lee, S., 2017. A unified approach to interpreting model predictions in 31st conference on neural information processing systems (nips 2017). CA, USA .
- Magen, E., Mishal, J., Zeldin, Y., Feldman, V., Kidon, M., Schlesinger, M., Sthoeger, Z., 2010. Increased mean platelet volume and c-reactive protein levels in patients with chronic urticaria with a positive autologous serum skin test. *The American Journal of the Medical Sciences* 339, 504–508. doi:10.1097/MAJ.0b013e3181db6ed5.
- Poddighe, D., 2019. Letter to the editor: The prevalence of chronic spontaneous urticaria (csu) in the pediatric population. *Journal of the American Academy of Dermatology* doi:10.1016/j.jaad.2019.07.068.
- Pozderac, I., Lugović-Mihić, L., Artuković, M., Stipić-Marković, A., Kuna, M., Ferček, I., 2020. Chronic inducible urticaria: classification and prominent features of physical and non-physical types. *Acta dermatovenerologica Alpina, Pannonica, et Adriatica* 29 3, 141–148. doi:10.15570/actaapa.2020.29.
- Rabelo-Filardi, R., Daltro-Oliveira, R., Campos, R., 2013. Parameters associated with chronic spontaneous urticaria duration and severity: A systematic review. *International Archives of Allergy and Immunology* 161, 197 – 204. doi:10.1159/000346896.
- Radonjic-Hoesli, S., Hofmeier, K., Micalotto, S., Schmid-Grendelmeier, P., Bircher, A., Simon, D., 2018. Urticaria and angioedema: an update on classification and pathogenesis. *Clinical Reviews in Allergy & Immunology* 54, 88–101. doi:10.1007/s12016-017-8628-1.
- Ring, J., Grosber, M., 2012. Urticaria: Attempts at classification. *Current Allergy and Asthma Reports* 12, 263–266. doi:10.1007/s11882-012-0275-2.
- SanchezBorges, M., Caballero-Fonseca, F., Capriles-Hulett, A., González-Aveledo, L., Maurer, M., 2017. Factors linked to disease severity and time to remission in patients with chronic spontaneous urticaria. *Journal of the European Academy of Dermatology and Venereology* 31. doi:10.1111/jdv.14221.
- Schettini, N., Corazza, M., Schenetti, C., Pacetti, L., Borghi, A., 2023. Urticaria: A narrative overview of differential diagnosis. *Biomedicines* 11. doi:10.3390/biomedicines11041096.
- Zuberbier, T., Latif, A.A.A., Abuzakouk, M., Aquilina, S., Asero, R., Baker, D., Ballmer-Weber, B., Bangert, C., Ben-Shoshan, M., Bernstein, J., Bindslev-Jensen, C., Brockow, K., Brzoza, Z., Neto, H.C.C., Church, M., Criado, P., Danilycheva, I., Dressler, C., Ensina, L., Fonacier, L., Gaskins, M., Gáspár, K., Gelincik, A., Giménez-Arnau, A., Godse, K., Gonçalo, M., Grattan, C., Grosber, M., Hamelmann, E., Hébert, J., Hide, M., Kaplan, A., Kapp, A., Kessel, A., Kocatürk, E., Kulthanian, K., Larenas-Linnemann, D., Lauferma, A., Leslie, T., Magerl, M., Makris, M., Meshkova, R., Metz, M., Micallef, D., Mortz, C., Nast, A., Oude-Elberink, H., Pawankar, R., Pigatto, P., Sisa, H.R., Gutiérrez, M.I.R., Saini, S., Schmid-Grendelmeier, P., Sekerel, B., Siebenhaar, F., Siiskonen, H., Soria, A., Staubach-Renz, P., Stingeni, L., Sussman, G., Szegedi, A., Thomsen, S., Vadasz, Z., Vestergaard, C., Wedi, B., Zhao, Z., Maurer, M., 2021. The international eaaci/ga2len/euroguiderm/apaaaci guideline for the definition, classification, diagnosis, and management of urticaria. *Allergy* 77, 734–766. doi:10.1111/all.15090.