computers

MDPI

Article

AI-Driven Clustering-Based Stratification of Allergic Patients Towards Smart Healthcare Systems in Southern Italy

Stefano Palazzo [1,2,*,†], Esra Hazar [3,†], Arife Uslu Gokceoglu [4], Giovanni Zambetta [5], Roberto Caldelli [1,6] and Claudio Loconsole [1,*]

[1] Department of Engineering and Sciences, Universitas Mercatorum, 00186 Rome, Italy; roberto.caldelli@unimercatorum.it
[2] "M. Albanesi" Allergy and Immunology Unit, 70126 Bari, Italy
[3] Pediatric Allergy and Immunology Unit, Department of Pediatrics, Alanya Alaaddin Keykubat University Faculty of Medicine, Antalya 07400, Türkiye; esra.hazar@alanya.edu.tr
[4] Department of Pediatrics, Alanya Alaaddin Keykubat University Faculty of Medicine, Antalya 07400, Türkiye; arife.gokceoglu@alanya.edu.tr
[5] Forensic Medicine, "F. Miulli", General Regional Hospital, 70021 Acquaviva delle Fonti, BA, Italy; vannizambetta@alice.it
[6] CNIT—National Interuniversity Consortium for Telecommunications, 50134 Florence, Italy
* Correspondence: stefano.palazzo@studenti.unimercatorum.it (S.P.); claudio.loconsole@unimercatorum.it (C.L.)
† These authors contributed equally to this work.

Abstract

A clustering analysis was conducted to identify distinct patient subgroups with White Blood Cells (WBC) count alongside Age and Total Immunoglobulin E (IgE) biomarkers. All data were obtained from a coordinated primary care network operating in Apulia (Southern Italy). We analyzed 300 patient records, performed preprocessing and exploratory data analysis, and then applied unsupervised clustering directly to the standardized three-variable feature space (Age, WBC, and Total IgE), followed by supervised validation steps. Several algorithms were applied for clustering. Among the evaluated methods, K-means and Spectral Clustering showed the most favorable internal validation profiles, based on Silhouette Score (SS), Calinski–Harabasz Index (CH), and Davies–Bouldin Index (DB). K-means achieved the best scores (SS = 0.406, CH = 190.00, DB = 0.900), closely followed by Spectral Clustering (SS = 0.398, CH = 182.57, DB = 0.936), outperforming Agglomerative Clustering (SS = 0.361, CH = 160.41, DB = 1.016) and Gaussian Mixture Models (SS = 0.233, CH = 103.89, DB = 1.289). Post-clustering ANOVA analyses indicated significant differences in WBC, age, and total IgE across the five consensus clusters. An evaluation of cluster internal separability occurred through the training of a Random Forest classifier to predict cluster membership. The results indicate internal cluster separability within the analyzed dataset, but more external verification and clinical evidence are necessary for validation. The research group established clinical descriptions along with suggested treatment plans and detected co-existing diseases to help validate model-based findings. A simplified cluster-informed clinical summary based on biomarker ranges was derived to support interpretation of the identified patient profiles. This integrated method preliminarily suggests that patient strata may be identified from routine clinical variables, while highlighting the importance of internal validation and clinical interpretability in clustering research.

Keywords: clustering analysis; clinical decision support; IgE; WBC; age; K-means; Random Forest classifier; supervised learning; unsupervised learning

Check for updates

Academic Editor: Rafiqul Chowdhury

Received: 26 February 2026
Revised: 26 April 2026
Accepted: 27 April 2026
Published: 7 May 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

1.1. Background and Current Drawbacks

Hundreds of millions of people worldwide suffer from allergic diseases such as allergic rhinitis, asthma, food allergies and atopic dermatitis, as their number continues to grow among children and adults [1]. Medical care becomes challenging because these diseases demonstrate various forms of clinical expression and treatment outcomes differ substantially between patients [2,3].

There is also consolidated evidence indicating that clinical, behavioral and social factors influence the outcomes and Quality of Life of allergic patients and their caregivers [4–7].

Current clinical practice requires identifying distinct clusters of allergic patients to enhance therapeutic care and achieve better treatment results along with modern personalized medicine goals [8].

The traditional approach for the clinical categorization of allergic patients is based on documented clinical signs as well as self-reported symptoms and biomarker measurements, often including specific Immunoglobulin E (IgE) assessments and eosinophil counts [9]. The current methods for disease identification yield inadequate results because they fail to delineate the complex structure of allergic pathologies. Two patients with equal IgE measurements may exhibit very different clinical patterns as their immune response patterns, age and comorbidities behave differently [10]. However, this irregularity leads to inadequate and excessive medical treatment, as well as patients receiving their diagnosis too late and suffering from inadequate treatment outcomes.

Allergic diseases are diagnosed and treated by different health practitioners such as general practice, allergy and pulmonology practitioners with the information evaluated by the health practitioner based on his/her expertise and training experience. Clinical categorization requires more objective and reproducible criteria, as current assessment and management practices can lead to inconsistent care.

It is necessary to stratify allergic patients according to a standardized, multimodal severity scale (integrating clinical and biological dimensions) in order to: (i) improve diagnostic assessment; (ii) personalize treatments; (iii) make research results comparable and reproducible [11–13].

1.2. Difficulties in Subjective Assessment and Reproducible Insights

Such subjectivity of testing and treatment choice has to be mentioned among the problems connected with the treatment and management of allergies. When physicians evaluate patients' symptoms using experience-based interpretations in combination with test results, these methods can lead to the development of biases and overload cognitive systems, leading to altered quality of care [14]. Because the accumulation of medical information becomes massive enough to include labor data analysis results and demographics, clinical histories alone will not be adequate for clinicians to uncover meaningful relationships between variables that could translate into diagnostic/prognostic value.

In the case of interaction of individual variables, the medical knowledge that we have acquired, as a result of our analysis of each variable, changes because interaction forms a new knowledge of patient immunological patterns. Medical practitioners should consider that pediatric patients with comparable high IgE levels have distinct risk characteristics than adult patients who show similar IgE concentrations.

A high White Blood Cells (WBC) count measurement, combined with moderate IgE levels, may suggest that an inflammatory process might occur even without any clinical manifestations. When traditional analysis methods or threshold classification methods are used, it is difficult to understand the comprehensive relationships between multiple factors.

Clinical medicine faces a reproducibility crisis rooted in biomedical research because non-reproducible stratification criteria lead practitioners to produce inconsistent treatment results [15].

A reliable patient classification method needs to produce equivalent subgroup identification across different healthcare providers and institutions with any patient sample. Clinicians should prioritize patient classification in allergy care because strategic choices such as immunotherapy initiation and avoidance strategies and referrals are directly affected by patient groups.

1.3. Engineering Value and Research Gap: Use of Unsupervised AI to Discover Unbiased Patient Profiles

Unsupervised machine learning has emerged as a fast-evolving AI technology that healthcare uses for detecting patterns and conducting classifications while making data-based decisions. Clustering algorithms demonstrate exceptional utility for clinical phenotyping because they use unguided methods to group patients through their input feature similarities without needing pre-defined outcome information [16].

The use of clustering methods has proven effective across different medical fields such as oncology and cardiology and infectious disease studies because they help identify new disease types and better forecast patient risks [17,18].

Research into unsupervised learning applications in allergy and immunology medicine remains limited even though it could generate new insights into immunophenotypes and personalized treatment recommendations [19].

Unsupervised clustering adds value by finding true patterns in multivariate data while eliminating human biases arising from assumed knowledge or predefined datasets. For successful detection of pathophysiological differences in allergic disease, it becomes crucial to overcome diagnostic ambiguity along with biomarker variability and symptom overlap [20].

The application of clustering algorithms to the healthcare variables age, WBC, and total IgE enables clinicians to detect previously imperceptible patient clusters using three medically relevant features. Studies that identify similar clusters using numerous clustering algorithms demonstrate increased result reliability through consensus cluster analysis according to Monti et al. (2003) [21].

1.4. Study Objectives and Novelty

This study investigates whether allergic patients can be meaningfully stratified in an unsupervised manner using age (years), White Blood Cell count (WBC, per mm^3), and total Immunoglobulin E (IgE, kU/L), while applying K-Means, Agglomerative Clustering (a hierarchical method), Gaussian Mixture Models (GMM), and Spectral Clustering to the same feature space.

The researchers chose these variables because they are of clinical significance and because they are accessible during provision of normal health services.

This deliberately restricted feature set reflects a parsimonious design. The aim was not to build a comprehensive etiological model of allergy, but to test whether routine, low-cost, clinically relevant variables could yield a meaningful and reproducible stratification. Age, WBC, and total IgE were chosen because they are widely available in outpatient care, easily obtainable in a screening-oriented setting, and easily transferable to ambulatory and EMR-based workflows. To the best of our knowledge, this minimal clustering-based approach has not previously been reported in allergic patients.

The variable age serves as a measure for immune system maturity while WBC count represents inflammatory status and IgE functions as the key marker for allergic factors. Together, these parameters provide basic clinical and immunological information about the patients.

To enhance reliability, several designs were employed such as the use of a consensus clustering methodology in which labels were reconciled using Hungarian Matching symmetry and at least three algorithms were needed to reach a consensus over the cluster outcome before that result was adopted [21–23]. This method allowed the science community to identify clusters of resilient patients with a selection of members that showed very low reliance on particular analysis systems.

The main objective of this work is to assess whether unsupervised clustering can identify clinically meaningful patient groupings, while the adoption of multiple clustering methods is intended only to support the robustness, reproducibility, and clinical interpretability of the resulting groups.

A simplified decision-support guide was derived from the most clinically stable clusters for potential clinical use. This manual guide allows doctors to access custom optimum therapeutic plans by using reference range data, e.g., interquartile ranges to assign patients into a cluster.

Furthermore, this study aims to develop an outpatient risk stratification framework for Southern Italy, using a regional primary-care cohort (Apulia) and aiming at integration into local workflows.

In line with these objectives, the remainder of the paper is organized as follows. Section 2, entitled "Study Design and Analytical Framework Overview", outlines the study aims, healthcare implications, and overall methodological workflow adopted in this work. Section 3, "Materials and Methods", describes the dataset, preprocessing steps, exploratory data analysis, clustering methods, consensus strategy, and validation procedures. Section 4, "Results and Discussion: Findings and Clinical Stratification", reports the main findings of the analysis, including the identification and characterization of the patient clusters, their statistical and clinical interpretation, and the proposed simplified decision-support guide. Section 5, "Conclusions and Clinical Implications", summarizes the main implications of the findings, discusses the methodological limitations of the study, and highlights future research directions.

2. Study Design and Analytical Framework Overview

2.1. Study Aims and Healthcare Implications

The following points summarize the specific objectives of this study and how each of them can be translated into routine outpatient care.

1. Parsimonious and Interpretable Stratification: Demonstrate that three routine measures (age, white blood cells, and total IgE) yield reproducible, clinically interpretable, and useful patient subgroups for outpatient management in Southern Italy;
2. In an unsupervised context, it is necessary to guarantee both the stability of the clusters and their reliability in the absence of reference labels. To this end, we distinguish two aspects:
    - Internal consistency of the Results: Construct clusters via ensemble-consensus strategy merging four paradigms, label matching via the Hungarian algorithm, and assignment stability via 3/4 majority rule; explain the ultimate selection of k = 5 based on typical internal indices and clinical consistency of disparities between subgroups. Subsequently, assess the internal consistency of the subgroups.
    - Estimation of actual reliability without a gold standard: Quantify the reliability of partitions with label-free indicators: assignment consensus rate, bootstrap stability, prediction strength on half-half splits, and silhouette distribution (median and tails), offering an interpretable measure of separability.

3. Translation into clinical actions: Derive an outpatient decision guide that specifies the frequency of check-ups, when to repeat tests (IgE and white blood cells), which targeted investigations to prioritize, and referral criteria, designed for integration into the electronic medical record in Apulia region. In fact, the study explicitly aims to stratify the risk for outpatient services in Apulia in order to provide region-specific decision support.

A key strength of the proposed framework lies in its clinical interpretability. By relying exclusively on routinely collected variables already used in daily practice, the stratification facilitates integration into electronic medical record systems and supports more consistent decision-making across healthcare providers, potentially reducing inter-physician variability in follow-up intensity and referral decisions.

2.2. Methodological Workflow

All analyses were performed starting from a single-region cohort collected in primary care in Apulia.

Figure 1 presents a summary diagram of the study's analysis flow that linearly summarises all the main methodological phases, from data acquisition to final assessment.

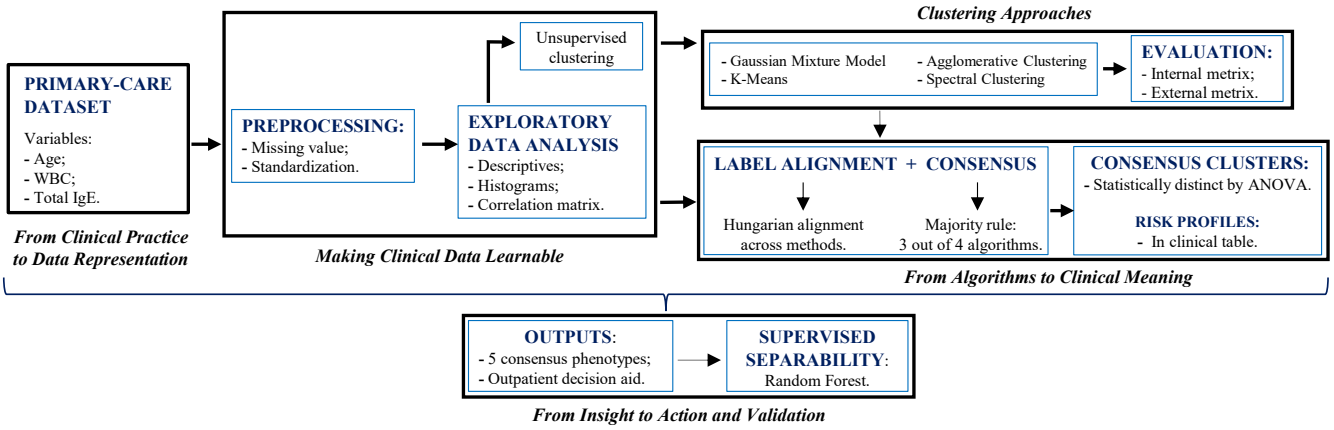


Figure 1. Overall implementation and methodological workflow of the proposed framework, illustrating the sequential analytical pipeline adopted in this study from data acquisition and preprocessing to clustering, consensus analysis, and final validation.

The Figure 1 describes dataset characteristics together with preprocessing steps followed by the utilized clustering models and consensus strategy methods alongside evaluation metrics for assessing clustering performance, including the subsequent translation of stable clusters into clinically interpretable phenotypes and decision-support outputs. Specifically, the Figure 1 aims to provide the reader with an overview of the order in which the different methodologies were applied, clarifying the logical sequence of the analyses conducted in this study. The block diagram highlights: (i) the acquisition of the real-world dataset and the selection of the clinical variables of interest (Age, WBC, total IgE); (ii) the preprocessing phases (checking for missing values and standardizing variables) and the Exploratory Data Analysis phase (descriptive statistics, histograms, and correlation matrix); (iii) the application of unsupervised clustering using four complementary approaches (Gaussian Mixture Model, Agglomerative, Spectral, and K-Means); (iv) the evaluation of the clustering solutions using internal validation metrics and inter-method agreement metrics; (v) label alignment (Hungarian matching) followed by a label-aligned majority-vote consensus procedure based on majority rule ($\geq$3/4 algorithms), which leads to the identification of five statistically distinct phenotypes (ANOVA) and translated into

clinical profiles and decision aids; (vi) the final verification of cluster separability through a supervised learning approach (Random Forest) for cluster membership prediction.

Against this background, this study describes the research design for applying unsupervised learning models in clinical data analysis with comprehensive methodology.

3. Materials and Methods

3.1. Dataset Description and Preprocessing

This research utilized data collected by medical professionals that contained real clinical information about a patient group that had allergic reactions. In particular, data were retrospectively gathered in a primary-care/ambulatory context within a network with coordinated participating general practitioners working in the Apulia region, who agreed to a shared export and undertook to submit in a single, fully de-identified Excel worksheet, for every patient, just the three variables used in this study. Data collection covered a 12-month period, from December 2023 to November 2024, and was based on information already present in the routine medical records of participating general practitioners. The source data were the routine medical records of the participating general practitioners. For study purposes, only the variables of interest were extracted and transferred into a fully de-identified dataset for analysis.

Specifically, all included subjects had a clinical diagnosis of allergic disease already established by physicians before data extraction. Eligible participants were adults ($\geq$18 years) with a physician-confirmed diagnosis of allergic disease and simultaneous availability of the three variables considered in this study (Age, WBC, and Total IgE). The proposed approach aims exclusively at stratification and decision support within a previously diagnosed population (in accordance with current guidelines and standards of care). Subjects were excluded only when one or more of these three variables were not concurrently available in the medical record at the time of data extraction. For each eligible subject, a single record was included in the final dataset. No direct identifiers were included in the research dataset, and only the variables necessary for the present analysis were transferred.

To obtain clinical specificity, patients in both groups who responded to food and inhalant allergens were investigated. The dual-allergen population enabled researchers to take note of the various different manifestations of the clinical conditions that were essential during the realization of the stratification measures.

In outpatient practice, allergic patients often exhibit mixed phenotypes with polysensitization to inhalant and food allergens. The inclusion of these heterogeneous presentations reflects the real-world medical context and reinforces the clinical relevance of our stratification approach.

In fact, the addition of allergens belonging to two distinct categories improved the diverse nature of the dataset which serves as an essential requirement for testing clustering algorithm stability.

The obtained clusters do not represent a subdivision based on the single allergen, but rather the result of a multivariate sensitivity analysis, aimed at capturing complex clinical and biological profiles.

The final cohort included n = 300 patients with confirmed allergic disease and active symptoms, undergoing diagnostic evaluation and/or undergoing treatment or clinical follow-up, with heterogeneous exposure and sensitization profiles and representative of the main allergic manifestations; a control group of healthy subjects was not included, since the aim of the study was to stratify allergic patients.

A previous examination indicated that sex did not alter the clustering structure so it is removed from further analysis. Sex did not appear in the last clustering analysis because exploratory findings showed weak discrimination ability. The literature demonstrates that

immune markers show sex-based variations which become secondary to age or inflammation status under specific circumstances [24–26]. The inclusion of sex as a variable added random elements which did not lead to better cluster discrimination. Future research should study sex-based group analyses to enhance subgroup stratification.

The analysis used clinically important data points which included:

- patient clinical evaluation took place during which researchers assessed their age through a variable named age (in years). A patient's age influences allergy pathophysiology because patient immunological responses commonly change regarding their maturity levels [27];
- a WBC count measurement expressed in cells per mm^3 (equivalently, cells per μL, since 1 mm^3 = 1 μL) allows medical professionals to track systemic inflammation and monitor immune system activities indirectly. Scientists have observed that high WBC numbers commonly link to both allergic responses and subsequent attacks of infection [28];
- the measurement unit of Total Immunoglobulin E (IgE) levels expresses these levels in kU/L while indicating allergic diseases as well as hypersensitivity. The clinical severity in allergic patients typically elevates when their Total IgE concentrations become higher [29].

The StandardScaler module from Scikit-learn Python library standardized all three variables (age, WBC, total IgE). Standardization transforms variables into a common scale with zero mean and unit variance to make them contribute equally to clustering procedures. The standardized matrix was used as input to all clustering algorithms and internal validation metrics, while descriptive tables and figures were computed and displayed in the original units (years, cells/μL, kU/L).

From a computational perspective, the proposed framework is intentionally lightweight and can be implemented in a standard data-analysis environment using a de-identified patient-level dataset including the variables considered in this study. From a clinical perspective, its practical use is based on information that is routinely available in outpatient care and is intended as a simple support for clinical interpretation, either as a reference table or within future EMR-based workflows.

3.1.1. Rationale for Variable Selection

The selection of these variables was driven by both clinical relevance and pragmatic considerations related to real-world primary care settings.

Consistent with the study objectives, the restriction to three variables was intentional and reflects the aim of developing a simple, accessible, and low-cost stratification framework for routine outpatient care. This study should therefore be interpreted as a preliminary, pragmatic proof-of-concept rather than as an attempt to capture the full biological complexity of allergic disease. The objective was not to build an exhaustive etiological model, but to test whether a minimal set of clinically meaningful and routinely available variables could generate reproducible and clinically interpretable patient profiles. Age, WBC, and total IgE were selected for their clinical relevance, routine availability, and suitability for a simple, transferable decision-support tool, including possible integration into EMR-based workflows. This deliberate minimalism also addresses a gap in the literature, where existing clustering-based stratification models generally rely on richer feature sets and are less applicable to real-world low-complexity settings.

In this context, Age is not considered merely as a demographic variable, but as a clinical modifier of immune response, disease chronicity, and cumulative allergen exposure, which may substantially influence the interpretation of similar IgE or WBC values across different age groups.

In particular, although eosinophil counts represent an important biomarker in allergic disease, total white blood cell count was deliberately selected to reflect a pragmatic primary-care perspective. Total WBC is universally available, less susceptible to pre-analytical variability, and routinely measured across outpatient settings, thereby enhancing the reproducibility and real-world applicability of the proposed stratification framework.

3.1.2. Ethical Considerations

This study was designed as a retrospective, observational, non-interventional analysis based exclusively on clinical data generated during routine medical care, in accordance with the standard of care and established clinical guidelines. No experimental interventions, randomization, or deviations from standard diagnostic or therapeutic procedures were performed.

In accordance with Italian national regulations and the Declaration of Helsinki of 1964 and its subsequent amendments, retrospective observational non-interventional studies that do not involve therapeutic or diagnostic modifications do not require prior approval by an Ethics Committee [30–32]. Specifically, under Italian law (Ministerial Decree of 30 November 2021, implementing EU Regulation 536/2014), such studies fall within the category of observational research and are subject to ethical assessment procedures that may result in exemption from formal authorization when no interventional elements are present.

All data were previously collected for clinical purposes and analyzed in aggregated form. No additional examinations, interventions, or modifications of patient management were introduced for research purposes. No individual-level data were published, and participant management was not influenced in any way. The shared data were fully and irreversibly anonymized at the source, in compliance with GDPR (EU Regulation 2016/679), including Recital 26, and relevant national implementing regulations (free of any personal and/or identifying references, ensuring complete confidentiality of the subjects involved).

Accordingly, ethical approval was considered waived under the above-mentioned conditions. In line with this, the Ethics Committee of Universitas Mercatorum (Rome, Italy) issued a determination of no grounds to proceed and did not undertake a formal ethical evaluation. More specifically, this determination was based on the retrospective, non-interventional, and fully anonymized nature of the study, for which no formal Ethics Committee review was required under the applicable regulatory framework.

Within this ethical and regulatory framework, the data were obtained through the collaboration of a network of general medical practices distributed across Apulia region, and therefore originate from a geographically limited area rather than a nationwide cohort.

3.2. Clustering Algorithms and Design Choices

Four unsupervised clustering algorithms were applied to identify distinct clinical profiles in allergic patients. They were chosen to cover complementary paradigms: centroid-based (K-means), hierarchical (Agglomerative), spectral graph–based (Spectral Clustering), and probabilistic modeling (GMM). The multiple clustering approaches contribute to creating consensus between distinct clustering logic systems.

In the following subsections, we briefly review the four clustering algorithms used in this study, highlighting for each the basic principle and the implementation choices adopted in the pipeline. This will motivate their application to clinical data and enhance the interpretability and reproducibility of the subsequent consensus analysis.

3.2.1. K-Means Clustering

K-Means breaks the input data into K separate clusters which do not overlap at all after classifying by features. K-Means clusters data points by finding centroids to minimize the sum of squares between the points in each identified group. The study set K to 5 clusters as a constant value for maintaining uniformity between all studied algorithms.

In practice, K-Means was applied using the *scikit-learn* library with Euclidean distance and Lloyd's algorithm. A fixed random seed (SEED = 42) was used to ensure reproducibility across runs. The corresponding implementation in code was `KMeans(n_clusters = 5, random_state = 42)` applied to the standardized feature matrix; in the adopted *scikit-learn* implementation, the remaining unspecified parameters (including `n_init`) were left at their default values of the software version used. Accordingly, the number of clusters was set to k = 5, and the algorithm was initialized multiple times to reduce sensitivity to centroid initialization (selecting the solution with the lowest within-cluster sum of squares).

3.2.2. Agglomerative Hierarchical Clustering

Each data point begins as its own cluster in this algorithm structure which uses an approach that works from the bottom up for cluster building. The algorithm continues group merging phases by implementing a linkage criterion which produces a cluster arrangement matching the target cluster number. According to the research this study depended on Ward's linkage technique for cluster minimization [33].

In practice, Agglomerative clustering (hierarchical clustering) was implemented using Ward's linkage criterion with Euclidean distance, as provided by the *AgglomerativeClustering* class in *scikit-learn*. The corresponding implementation in code was `AgglomerativeClustering(n_clusters = 5)` applied to the standardized data matrix. In the adopted *scikit-learn* implementation, this setting corresponds to Ward linkage and Euclidean distance, with the remaining unspecified parameters left at their default values. Consistently, the dendrogram was cut to obtain k = 5 clusters, in line with the other clustering strategies.

3.2.3. Gaussian Mixture Model

This algorithm views data as generated outputs from multiple Gaussian distributions that respectively map individual clusters. The covariance structure of data becomes a consideration in GMM clustering because it enables the algorithm to identify elliptical cluster shapes. The Expectation-Maximization algorithm serves to determine Gaussian component parameters according to Huang et al. [34].

In practice, the Gaussian Mixture Model was implemented using the *GaussianMixture* class from *scikit-learn*, with: (i) the number of components fixed to k = 5; (ii) full covariance matrices to allow elliptical clusters; (iii) parameters estimated via the Expectation–Maximization algorithm. The corresponding implementation in code was `GaussianMixture(n_components = 5, random_state = 42)` fitted on the standardized feature matrix; in the adopted *scikit-learn* implementation, this corresponds to the default setting `covariance_type = 'full'`. Since GMM is a probabilistic model, cluster assignment was based on posterior probabilities under the fitted Gaussian mixture rather than on a single explicit distance metric. Accordingly, a fixed random seed (SEED = 42) was used to ensure reproducibility.

3.2.4. Spectral Clustering

Clustering uses eigenvalues of similarity matrices to achieve dimension reduction as its first step toward clustering. The methodology proves beneficial when clustering non-convex data structures and requires fewer initialization requirements than K-Means clustering [33]. In practice, Spectral clustering was implemented using the *SpectralClustering* algorithm from *scikit-learn*, adopting a k-nearest-neighbors affinity graph to construct the similarity matrix. The corresponding implementation in code was `SpectralClustering(n_clusters = 5, affinity = 'nearest_neighbors', random_state = 42)` on the standardized feature matrix; therefore, the number of neighbors used to build the affinity graph (`n_neighbors`) was left at the default value of the

adopted *scikit-learn* version. Thus, similarity was defined through a nearest-neighbors affinity graph rather than through direct centroid-based Euclidean assignment.

The number of clusters was fixed to k = 5. And a fixed random seed (SEED = 42) was used to control the stochastic components of the algorithm.

3.2.5. Implementation Specifics and Reproducibility Settings

The full analysis pipeline was executed using Python (version 3.9).

More specifically, the analysis was implemented using Python 3.9.13 and the *pandas* (v2.2.3), *NumPy* (v2.0.2), *SciPy* (v1.13.1), and *scikit-learn* (v1.6.0) libraries. The analysis was implemented using the *pandas*, *NumPy*, *SciPy*, and *scikit-learn* Python libraries. Before clustering, Age, WBC, and Total IgE were transformed using `StandardScaler()`, so that all clustering procedures were applied to the same z-score standardized matrix with zero mean and unit variance for each feature.

A thorough evaluation process requires setting common conditions among the various clustering algorithms. The number of clusters was set to k = 5 for all clustering models, based on a combination of methodological criteria and clinical considerations. A preliminary analysis included the use of the elbow method and the silhouette score trend as k varied. This analysis revealed that the 5-cluster solution represented an adequate compromise between intra-cluster compactness and inter-cluster separation, avoiding both excessive profile aggregation and unstable fragmentation. This configuration was also discussed and shared with the allergists involved, who confirmed its clinical consistency and compatibility with a pragmatic interpretation of the profiles in the outpatient setting.

The implementation used a shared random seed (SEED = 42) for the purpose of achieving replicable results. More specifically, the code set both the Python built-in random generator and the NumPy random generator to the same seed value (`random.seed(42)` and `np.random.seed(42)`), and the same seed was passed to the stochastic *scikit-learn* models whenever applicable. All features (Age, WBC, and Total IgE) were standardized to zero mean and unit variance prior to fitting. The combination of Euclidean distance together with Ward's linkage method operated as the linkages and distance metrics for hierarchical clustering. More specifically, for K-means, clustering was performed in the standardized feature space under Euclidean geometry, whereas Spectral Clustering relied on a nearest-neighbors affinity graph. For GMM, the assignment mechanism was probabilistic and derived from the fitted mixture model rather than from a single predefined distance function.

Spectral Clustering relied on a k-nearest-neighbors affinity graph; K-Means used standard initialization with multiple starts; and the Gaussian Mixture Model allowed full covariance among features.

For cross-method comparison and consensus, cluster labels were aligned across algorithms using the Hungarian matching procedure to avoid label switching (by solving a maximum-overlap assignment problem on the confusion matrix between clusterings, as implemented in *scipy.optimize.linear_sum_assignment*). In the implemented pipeline, K-means labels were used as the reference labeling scheme, and the labels obtained from Agglomerative clustering, GMM, and Spectral clustering were remapped accordingly before pairwise comparison and consensus voting.

Unless otherwise specified, default settings of the adopted *scikit-learn* implementations were used for all algorithms. Accordingly, the exact behavior of any parameter not explicitly reported in the manuscript followed the defaults of the software environment used for the analysis, which should be considered when reproducing the pipeline in a different *scikit-learn* version.

From a computational perspective, the proposed framework has modest system requirements and does not require high-performance computing resources or dedicated GPU acceleration. All analyses were performed on a standard 64-bit personal computer equipped with a 12th Gen Intel Core i7-12700H processor (2.70 GHz) and 16 GB RAM, using Python 3.9.13 together with the pandas, NumPy, SciPy, and scikit-learn libraries. Under these conditions, the complete analytical workflow—including preprocessing, clustering, consensus analysis, and validation—was completed in less than one minute on the full dataset. These results indicate that the framework is computationally lightweight and feasible for implementation in ordinary modern data-analysis settings.

3.3. Evaluation Metrics and Validation of Clustering Results

Clustering algorithm evaluation was based on both internal validation metrics, inter-method agreement metrics, and statistical testing.

Given the unsupervised nature of the problem, metrics such as ARI, NMI, V-measure, and FMI were used to quantify agreement between clustering solutions (inter-method consistency), rather than comparison with a reference partition.

Since the study was fully unsupervised and no clinician-assigned external risk classes were available, the five profiles were identified in a data-driven manner rather than compared against a pre-existing clinical taxonomy. Accordingly, no external validation in the strict sense was possible.

Accordingly, multiple evaluation criteria were adopted to provide a comprehensive assessment of how well the clustering models identify meaningful data structures. In this context, the clinical relevance of the groupings was evaluated by jointly considering age, white blood cell count, and total IgE levels.

Within this framework, the clustering structure received evaluation through internal metrics which evaluated the data features without dependence on external labels or references. All internal validation metrics were computed on the same standardized feature matrix used for clustering, thereby ensuring methodological consistency between model fitting and cluster evaluation. The Silhouette Score served as the initial selected internal validation measure. More specifically, in the present implementation, the Silhouette score was computed using the default Euclidean metric of the adopted *scikit-learn* function. The Calinski–Harabasz and Davies–Bouldin indices were computed on the same standardized feature matrix using the corresponding default *scikit-learn* implementations. Silhouette Score calculates the difference between intra-cluster point similarity and inter-cluster point similarity. The score spans between -1 and 1 with superior matching to cluster members signifying increased score numbers [35]. This metric helps detect both the internal strength of clusters together with the distance between clusters.

The Calinski-Harabasz Index (CH) functioned as the second method to measure cluster analysis by comparing the ratio of inter-cluster to intra-cluster variances. Clustering results with higher CH scores indicate dense clusters which are properly separated from each other [35]. The Davies-Bouldin Index (DB) analyzes the average similarity between each cluster and its most similar cluster to obtain a value. A lower DB score indicates superior cluster performance because it shows compact and separated cluster arrangements.

An agreement-based evaluation measured the clustering results against a consensus clustering solution. The Adjusted Rand Index (ARI) serves to measure similarities between two clustering solutions after factoring in random element distribution patterns [36]. Excellent agreement exists when ARI measurements approach 1.

The Normalized Mutual Information (NMI) evaluates cluster assignment dependence on consensus clusters to determine clustering quality through its higher value assessment [37]. The V-Measure tool achieves evaluation results through balanced assessment

by combining both homogeneity measures (cluster exclusivity to specific classes) and completeness measures (proper class assignment). The Fowlkes-Mallows Index (FMI) combines precision and recall through their geometric mean to evaluate clustering accuracy in grouping same-type instances [38].

A statistical testing approach was then used to determine whether clinical variables differed significantly across the identified clusters. ANOVA testing evaluated the statistical significance determining whether age values plus WBC counts and Total IgE levels showed different mean distributions between cluster groups. ANOVA stands as a durable technique for detecting substantial statistical differences across various groups according to Handschuh et al [39]. In the present study, it was used as a post-clustering comparative analysis. More specifically, one-way ANOVA was used to test whether the consensus clusters differed in mean WBC, Age, and Total IgE values. Its main assumptions are independence of observations, approximate within-group normality, and homogeneity of variances across groups. Independence was supported by the inclusion of one record per patient. Since this analysis was intended as an exploratory comparison after clustering, ANOVA results were interpreted together with descriptive statistics and cluster visualizations. When ANOVA results are significant, they provide descriptive evidence that the clusters differ statistically on the variables considered.

3.4. Consensus Clustering Strategy

In this study, the term "consensus clustering" refers to a label-aligned majority-vote agreement procedure rather than to a formal co-association-matrix ensemble method.

Within this framework, the consensus approach was used to improve both the stability and the medical significance of the discovered clusters, since different clustering algorithms showed variable outcomes. The same data set leads to various cluster groupings when different clustering models operate on it primarily because of dissimilar algorithmic assumptions as well as distance measurement choices and search optimization methods. Medical institutions need reliable clustering results so the sole utilization of a single clustering model creates uncertainty while increasing possibility of faulty outcomes. The selection of consensus clustering occurred because it merges diverse model strengths to produce clinically applicable and more robust patient profiles [40].

The clustering approach considered age, WBC and total IgE because of their clinical significance and their relationship results from the exploratory analysis. The three variables offer unique biological information about age which reflects immunological aging and WBC which detects inflammatory states and Total IgE which measures immunological responsiveness. These features create new distribution patterns between clusters when analyzed together that are not detectable if used individually. Each variable used independently produced preliminary models that generated clusters with lower clinical value and cohesion.

The main limitation in consensus clustering arises when computational algorithms offer random numerical terminology for cluster groupings. The assignment of Cluster 1 from K-Means to Cluster 3 in Agglomerative Clustering creates problems for direct model comparison. The Hungarian Algorithm provided a solution to this issue. The Hungarian Algorithm gets the assignment problem solved efficiently because it guarantees that the labels on clusters are the best possible ones based on the maximum overlaps of solutions [41]. Proper alignment of group entities in the context of clustering solutions was a significant starting point in the determination of consensus strategy during this step.

The algorithm discovered a stable cluster when the four analysis methods used on the data all labeled the same cluster tag on a point in the dataset. Such a voting system where the majority wins makes sense in the analysis process. This process excludes those data that

cannot be classified compatibly because it demands majority of the analysis models to agree on them. These latter clusters consist of the patients whose merits of the methodological component of this method belong to more explicable, more plausible phenotypes.

Two main arguments justify the use of consensus clustering methods in this case: (i) This method enhances model stability because it reduces the effects of biases that are reflected in individual methods. Clustering with a variety of models allows developing the stable and accurate clustering since no algorithm can provide the perfect view of the data structure alone; (ii) it improves clinical usability. Several models that prove the same cluster groups enhance the probability of establishing the fact that in the data there exists actual clinically significant difference between patients. Agreeable created profiles will be more likely to be included in a decision support tool, that would increase the accuracy and personalization of patient care [36].

After label alignment using Hungarian matching, the majority rule defined consensus cluster membership when a sample was assigned to the same cluster by at least three of the four clustering algorithms. Operationally, for each patient, the final consensus label was assigned only when at least three of the four aligned clustering methods returned the same cluster identifier; otherwise, the case was labeled as unassigned (-1) and excluded from downstream consensus-based analyses. The threshold has been so fine-tuned that it maintains good scores and is cluster-sensitive. Since unanimity trims away too many samples and is willing to tolerate two out of four samples as potentially containing noise, both less lenient and more lenient thresholds can be problematic. To turn the consensus ensemble approach into a stability reinforcing method, the introduction of this three-of-four rule can inhibit the consensus ensemble approach, by eliminating samples that are unnecessary.

3.5. Internal Proxy Assessment of Cluster Separability Using Supervised Learning

Cluster validation used a Random Forest (RF) classifier because it showed strong performance on structured medical data as well as nonlinear capability and anti-overfitting properties from its decision tree ensemble [42,43]. Random Forest provides interpretable feature importance metrics without needing extensive preprocessing and it can handle data without requiring feature scaling [44] while alternatives such as Support Vector Machine (SVM) [45] and logistic regression [46,47] and k-Nearest Neighbors (k-NN) [48] do not share these features. RF was selected for model implementation because it demonstrated proven effectiveness on analogous clinical stratification problems and accommodates easy pipeline integration during exploratory analysis.

The Random Forest classifier was implemented using scikit-learn with the following parameters: number of trees = 100 (which is often used as a default to guarantee stability in predictions) [49,50], maximum tree depth = 8, and minimum samples per leaf = 2, both of which correspond to a trade-off between model complexity and the risk of overfitting [51,52].

To minimize the risk of overfitting in the Random Forest classifier, a 70/30 train-test split was employed, and model performance was reported only on the held-out test set.

While Random Forest is inherently robust to overfitting due to its use of bootstrapped aggregation and random feature selection, hyperparameter tuning was also conducted to avoid excessive depth or complexity. In fact, hyperparameter tuning was performed using 5-fold cross-validation on the training set, optimizing for F1-score as recommended for enhancing performance beyond defaults [43,49,53]. Feature importance was calculated using Gini impurity reduction [54,55].

In addition, cluster separability was further assessed through Stratified 5-Fold Cross-Validation of the Random Forest classifier, with Macro F1-score reported alongside accuracy, and by means of the out-of-bag estimate.

4. Results and Discussion: Findings and Clinical Stratification

For clarity, the results are presented in four sequential steps: (i) data quality and exploratory analysis; (ii) comparison of clustering algorithms; (iii) derivation and interpretation of consensus clusters; and (iv) internal proxy assessment of cluster separability using supervised learning.

4.1. Data Preprocessing

At a high level, this clinical research analyzed latent group structures in a 300-subject dataset that contained White Blood Cell (WBC) count measurements and age values and Total Immunoglobulin E levels. Specifically, the research intent involved examining data clusters through automated clustering techniques that based on visual aids and supervised test methods to derive clinical assessment models from patient characteristics.

In this context, and as a prerequisite for the analyses, any successful model development depends heavily on preprocessing data before analysis. Accordingly, in this project, before meaningful analysis began the dataset received an inspection for missing data along with outliers as well as inconsistencies.

4.1.1. Handling Missing Values

The examination showed missing values were absent from all entries in the features WBC, age and total IgE. Consistently, the preliminary data preprocessing step revealed excellent data quality since every value existed for the three variables. Therefore, since all the WBC, age and total IgE values were present authoring any data cleaning procedures for missing data was not needed. Overall, the dataset contained 300 complete samples alongside the 3 features as evidenced by its observed (300, 3) shape.

4.1.2. Outlier Screening

Inspecting boxplots of individual features allowed the identification of any outliers. The boxplots indicated these results as the observation ranges shown below and also that can be seen in Figure 2.

- WBC: 0 missing values;
- Age: 0 missing values;
- Total IgE: 0 missing values;

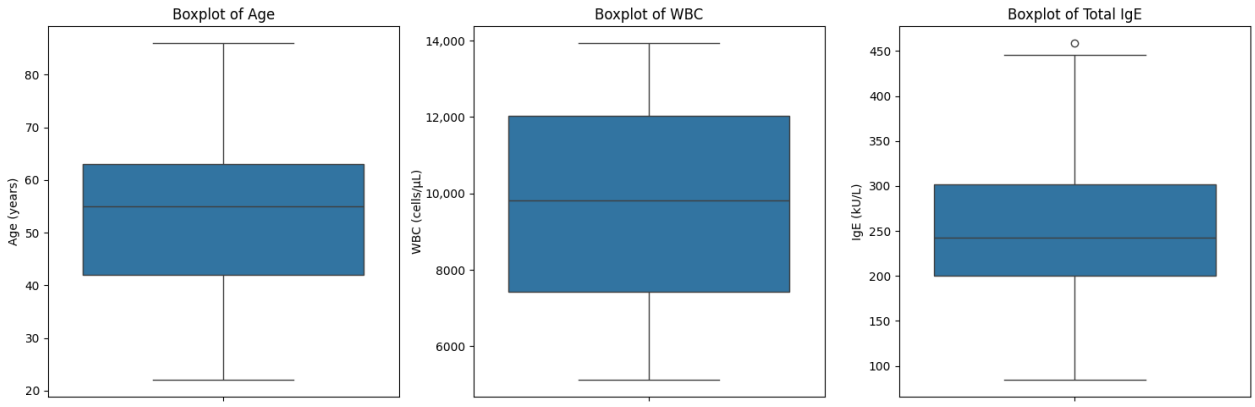


Figure 2. Outlier detection for Age, WBC, and total IgE using boxplots, showing the distribution and variability of each variable.

Some values lay beyond the whiskers, likely reflecting normal biological variability; therefore, all observations were retained. All data points were retained in the analysis, since the sample used in the study was not distorted in any way due to the outlier investigation process.

The boxplots analysis found distinct ranges that spanned from ≈5000 to ≈14,000 WBC cells/µL together with Total IgE levels from ≈80 to ≈460 kU/L and age values between ≈20 and ≈86 years.

4.2. Exploratory Data Analysis

The Exploratory Data Analysis (EDA) follows data preprocessing and in this stage, the first step would be Descriptive statistics of the features whose summary has already been mentioned in Table 1. Statistical measures used in analysis showed that there was a large difference between the values of the data points. As WBC results were between 5114 and 13,940, it appears that the patient sample had a variety of clinical features. The sample population was also broad with early adult participants (22 years old) at one end and late elderly participants (86 years old) at the other end (which demonstrates a broad participant demographic range). The fact that the Total IgE measurement outcomes vary greatly supported the theory that there are large differences in the immune system responses of each individual patient.

Table 1 summarizes the descriptive statistics in original clinical units to improve readability and facilitate interpretation.

Table 1. Descriptive statistics of the WBC, Age, and total IgE features, showing count, mean, median, standard deviation, percentiles, minimum, and maximum values.

Statistic	WBC	Age	Total IgE
Count	300	300	300
Mean	9749.54	53.66	251.22
Std. Dev.	2506.91	13.53	72.21
Minimum	5113.87	22.00	84.25
25th Percentile	7427.31	42.00	199.92
Median (50%)	9819.78	55.00	242.60
75th Percentile	12,035.65	63.00	301.70
Maximum	13,939.69	86.00	459.01

4.2.1. Feature Distributions

To obtain an idea of the distribution of the features histograms of the features were plotted (age, WBC, and total IgE) to visualise the distributions and Figure 3 shows the distribution of the features where age follows a distribution with centre 55 years, the WBC values are marginally right skewed and the total IgE levels also have a slight right skewness.

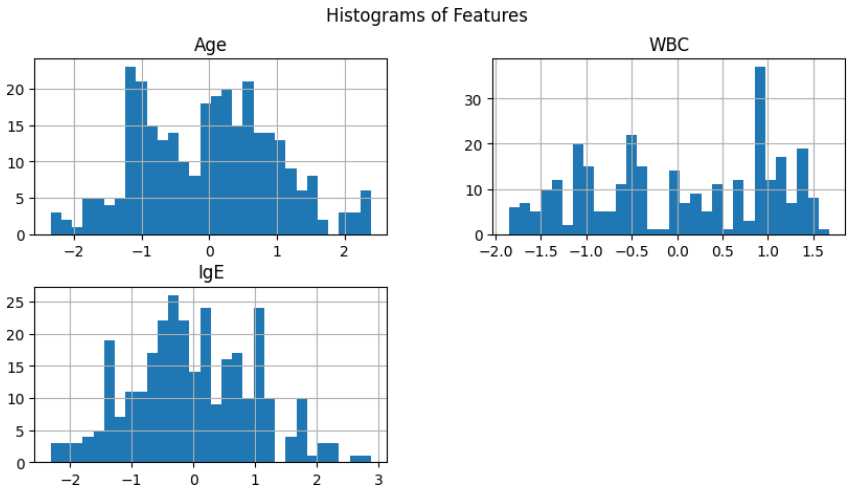


Figure 3. Histograms showing the ranges and empirical distributions of Age, White Blood Cell count, and total IgE in the study population. The x-axis represents the values of each variable, while the y-axis indicates the frequency of observations.

4.2.2. Correlation Analysis

The correlation matrix revealed associations among the features, as shown in Figure 4, where labeled variables and correlation coefficients facilitate interpretation.

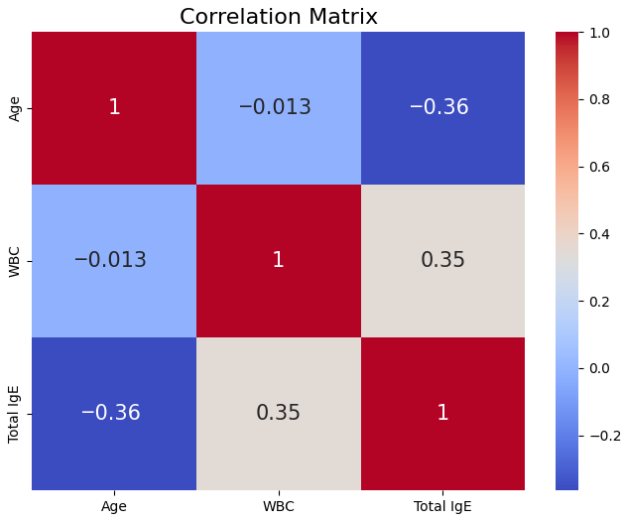


Figure 4. Heatmap of the correlation matrix between Age, WBC, and total IgE. Positive values indicate that variables tend to increase together, whereas negative values indicate that one variable decreases as the other increases.

Weak to moderate correlations were observed between the clinical parameters. In particular, a moderate negative relationship was observed between Total IgE and age.

A moderate negative correlation was observed between age and total IgE levels ($r = -0.36$), suggesting that higher IgE levels tend to be more frequent in younger individuals, while older patients generally present lower IgE concentrations. This pattern may reflect age-related changes in immune reactivity or differences in allergic disease expression across age groups. In contrast, white blood cell count showed almost no correlation with age ($r = -0.013$), while a moderate positive correlation was observed between WBC and total IgE levels ($r = 0.35$), indicating that higher IgE levels may be associated with increased systemic immune activity.

These relationships support the inclusion of the three variables in the clustering algorithm, as they capture partially distinct but complementary dimensions of clinical variation. Thus, the exploratory analysis provided a coherent basis for the subsequent multivariate clustering step.

4.3. Clustering Analysis and Comparison of Techniques

As introduced in the previous section, the number of clusters was set to $k = 5$, based both on methodological considerations and on clinical input from the allergists involved in the study. This choice will be further supported by quantitative evaluation metrics presented in the following sections.

As a preliminary qualitative check of cross-method consistency, Figure 5 reports the 3D clusterings obtained by each algorithm with aligned labels and consistent colors. This figure illustrates the degree of agreement that motivates the subsequent consensus analysis.

Clustering methods were compared using Adjusted Rand Index (ARI), Normalized Mutual Information (NMI), V-Measure, and Fowlkes–Mallows Index (FMI). These metrics were used to assess agreement between clustering solutions, not external validity against independent labels. For interpretative consistency across methods and for the subsequent majority-vote consensus procedure, cluster labels were aligned using Hungarian matching. Agreement metrics were then used to quantify similarity between the resulting partitions.

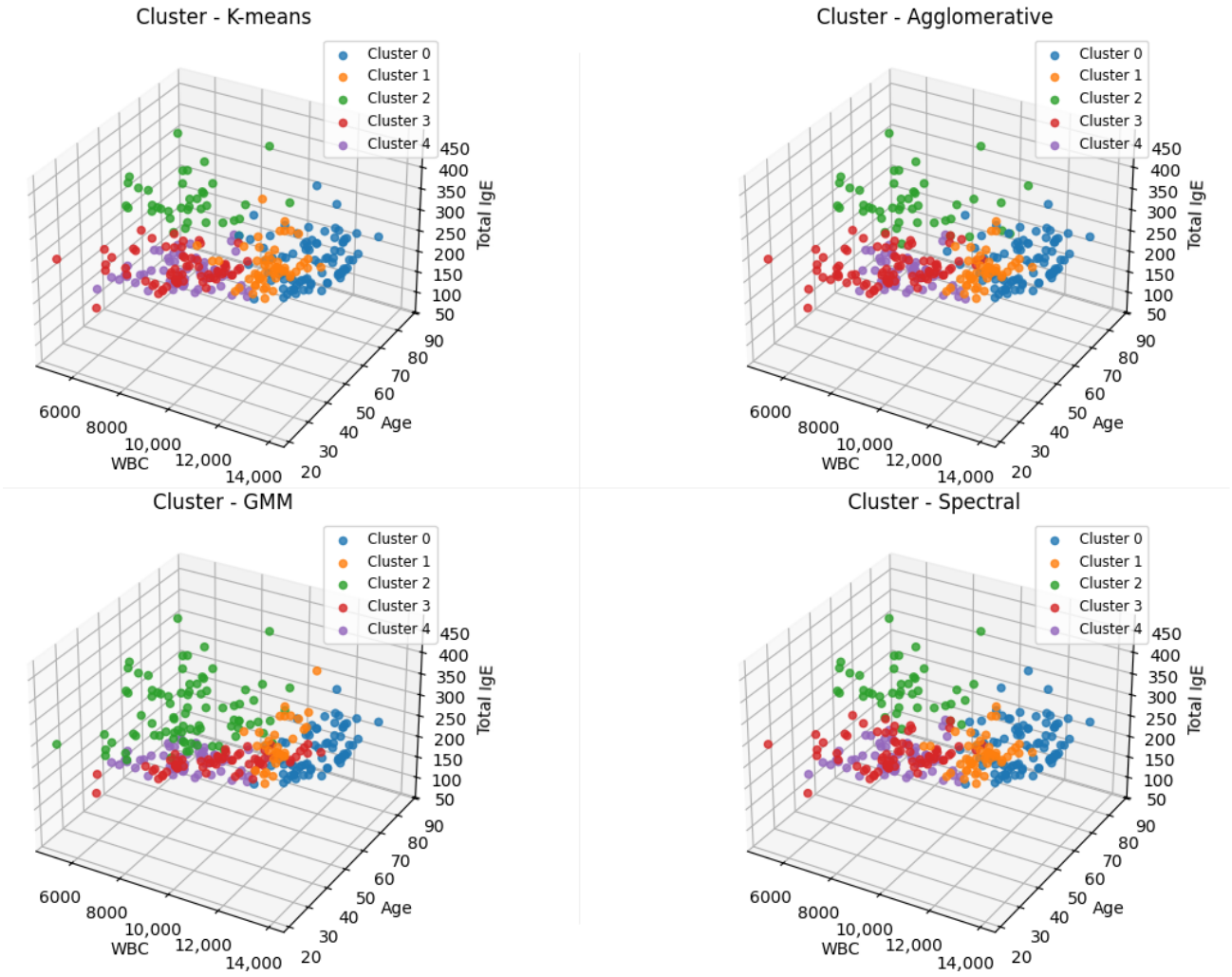


Figure 5. Comparative 3D plots of the clustering solutions produced by K-means, Agglomerative clustering, Gaussian Mixture Model, and Spectral Clustering, using a consistent color scheme to facilitate visual comparison and highlight agreement among cluster assignments.

Results are summarized in Table 2 and also a comparative 3D plot with consistent color for different clustering techniques is shown in Figure 5.

Table 2. Comparison of clustering methods using ARI, NMI, V-Measure, and FMI, showing the agreement scores between K-means, Agglomerative clustering, Spectral Clustering, and Gaussian Mixture Model.

Method 1	Method 2	ARI	NMI	V-Measure	FMI
K-means	Agglomerative	0.791	0.787	0.787	0.835
K-means	GMM	0.526	0.613	0.613	0.626
K-means	Spectral	0.899	0.883	0.883	0.920
Agglomerative	GMM	0.495	0.588	0.588	0.603
Agglomerative	Spectral	0.823	0.824	0.824	0.860
GMM	Spectral	0.543	0.630	0.630	0.639

Different clustering algorithms ran simultaneously to determine sustainable classification results. Spectral Clustering and K-means showed a notable accomplishment in clustering performance through their evaluation metrics which resulted in Adjusted Rand Index scores of 0.899 and 0.791 between them, further confirmed by complementary metrics such as NMI, V-Measure, and FMI, which collectively highlighted these two as the most stable clustering approaches among the evaluated methods.

In this analysis, the values of NMI and V-measure coincide due to their closely related formulations based on normalized mutual information. High values across ARI, NMI, V-measure, and FMI indicate a strong agreement between the clustering solutions.

Internal Metrics for Each Clustering Method

Internal metrics such as Silhouette Score, Calinski-Harabasz Index, and Davies-Bouldin Score were computed and are reported in Table 3.

Among the four algorithms, K-means showed the best overall internal validation profile, with the highest Silhouette Score and Calinski–Harabasz Index and the lowest Davies–Bouldin score. Spectral Clustering showed very similar performance, supporting the robustness of the main partitioning structure.

Table 3. Silhouette Score, Calinski–Harabasz Index, and Davies–Bouldin Score for K-means, Agglomerative clustering, Gaussian Mixture Model, and Spectral Clustering.

Method	Silhouette Score	Calinski–Harabasz Index	Davies–Bouldin Score
K-means	0.406	190.00	0.900
Agglomerative	0.361	160.41	1.016
GMM	0.233	103.89	1.289
Spectral	0.398	182.57	0.936

These results suggest comparatively favorable clustering performance, which is relevant in medical applications requiring reliable and interpretable groupings.

The Silhouette Score achieved 0.406 for K-means, as shown in Figure 6, which was plotted between the Silhouette score and number of clusters. Spectral clustering yielded a value of 0.398. Along with the Calinski-Harabasz Index and Davies-Bouldin Score results, these metrics supported the choice of k = 5 as a pragmatic number of clusters.

Higher Silhouette and Calinski-Harabasz values together with lower Davies-Bouldin scores indicate better cluster compactness and separation.

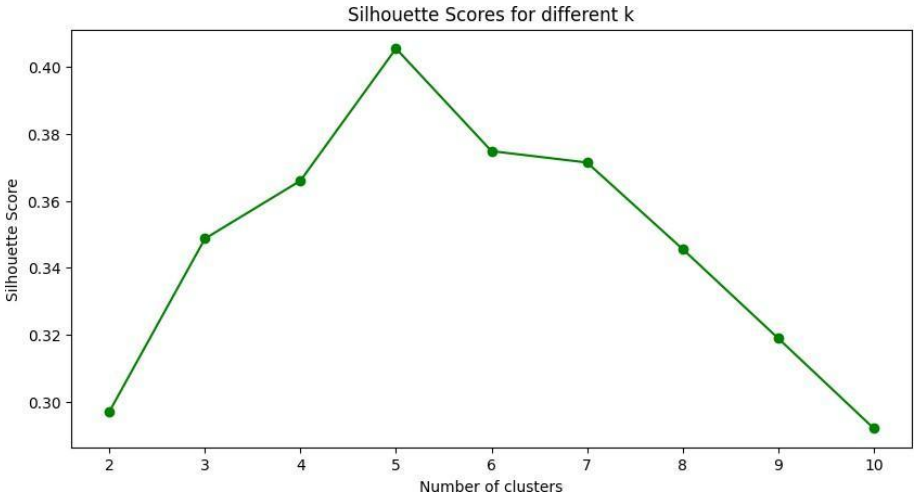


Figure 6. Curve showing the relationship between Silhouette Score and the number of clusters, reporting the Silhouette Score for different values of k.

This choice was also supported by the Elbow method analysis (Figure 7), where the within-cluster sum of squares (WCSS) curve exhibits a clear "knee" around k = 5, after which further increases in k yield only marginal reductions in WCSS, indicating diminishing returns in compactness.

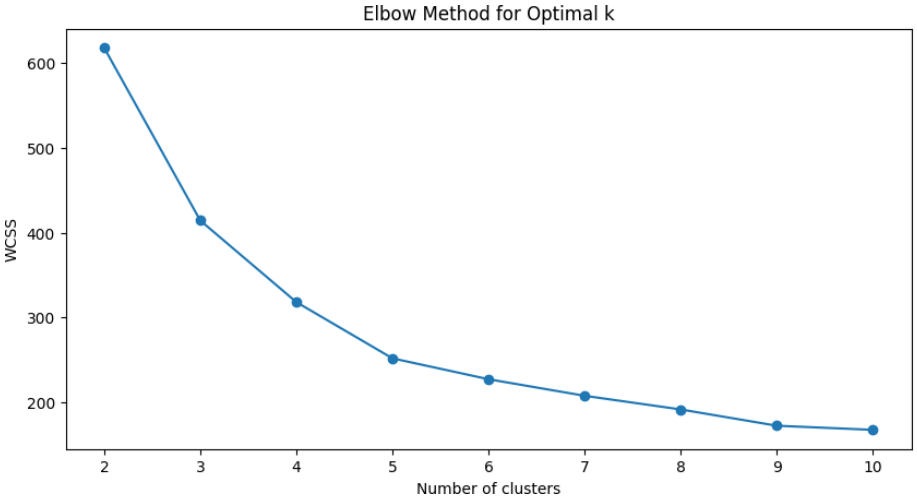


Figure 7. Elbow method curve for the selection of the optimal number of clusters, showing within-cluster sum of squares WCSS values for different numbers of clusters.

The outcome of Agglomerative clustering was satisfactory, but Gaussian Mixture Models (GMM) showed minor deficits in both inter-method agreement and internal verification metrics, given the imperfect fit of the Gaussian distribution to clinical data.

4.4. Consensus Clustering Analysis

A 3D plot of consensus clusters (with colors according to K-means clustering) is shown in Figure 8. Applying the 3-out-of-4 majority rule resulted in $N_{\text{stable}} = 287$ patients assigned to concordant cluster labels and $N_{\text{unassigned}} = 13$ patients labeled as unassigned, out of the total cohort of 300 patients. Therefore, most patients showed stable cluster membership across the different algorithms.

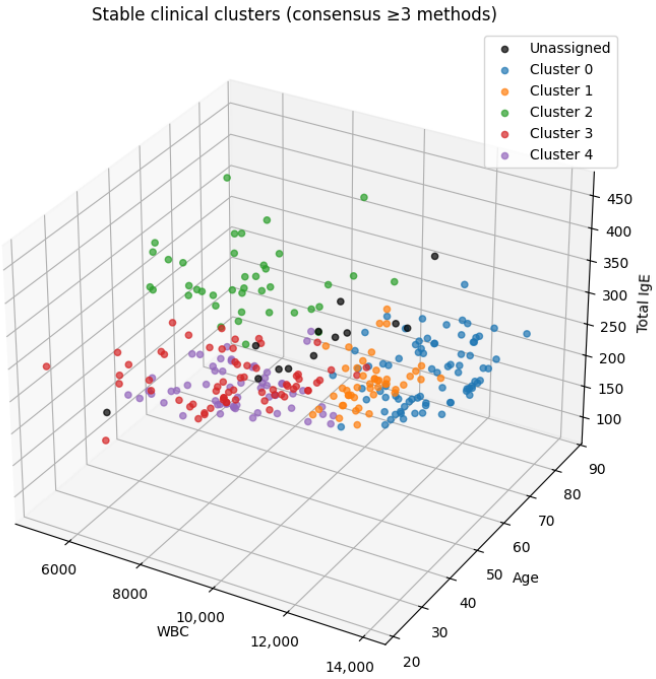


Figure 8. A 3D plot of consensus clusters (with colors according to K-means clustering) to show stable clinical clusters based on consensus from at least 3 methods. Cluster membership was assigned when at least three of four clustering algorithms agreed; patients not reaching this agreement were labeled as *unassigned* (shown in black), indicating inconsistent cluster assignments across methods.

The 13 subjects who could not be assigned according to this consensus rule were excluded from downstream analyses to preserve cluster stability. This choice, which was shared by the physicians, reflects their limited relevance for defining reproducible profiles.

In the Figure 8, it is clearly shown that clusters were grouped from cluster 0 to cluster 4. The clusters are mainly concentrated regions having WBC count as 8000–9000, an age group lying between 70–80, and whose Total IgE levels are less than 100. The same type of conclusion is found in Figure 3 using Boxplots of the stable clusters, which illustrate feature distributions within clusters of different groups.

4.4.1. Clinical Profile of the Consensus Clusters

The clinical profiles (mean, standard deviation, minimum, maximum, and count) for each consensus cluster (voted by at least three methods) are presented in Table 4. These descriptive statistics provide the basis for the clinical interpretation of each cluster.

Table 4. Clinical profile of the consensus clusters, reporting mean and standard deviation values for WBC, Age, and total IgE.

Cluster	WBC (Mean)	Age (Mean)	Total IgE (Mean)	WBC (Std)	Age (Std)	Total IgE (Std)
0	12,067.38	65.50	230.85	982.07	8.73	34.61
1	12,521.79	38.02	332.69	589.05	6.24	36.29
2	7323.33	61.25	316.75	1368.22	9.10	58.40
3	8584.14	41.60	256.26	1490.74	6.14	34.10
4	7313.74	58.70	152.12	1265.03	7.70	29.02

In addition, the distributions of Age, WBC, and Total IgE within each consensus cluster are visually summarized in Figure 9 using boxplots, highlighting intra-cluster variability, median trends, and potential overlap among groups.

In practical terms, Cluster 1 represents the profile with the highest combined WBC and IgE values in younger patients, whereas Cluster 4 represents the lowest-priority profile, with lower WBC and IgE values. Cluster 0 differs from Cluster 1 mainly by older age and lower IgE despite high WBC, suggesting a more inflammation-oriented profile.

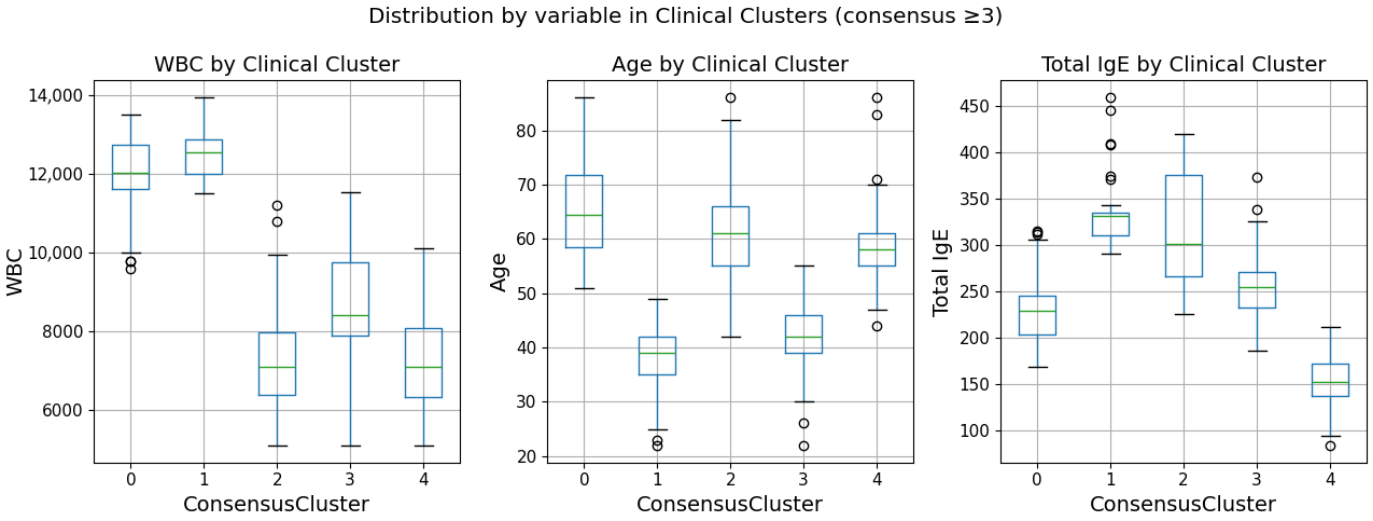


Figure 9. Distribution of Age, WBC, and total IgE within the clinical clusters identified by consensus of at least three methods, shown using boxplots.

4.4.2. ANOVA Analysis

To verify the significance of differences between clusters, ANOVA was conducted for each feature. ANOVA was used as a post-clustering descriptive analysis. Since it is applied to the same variables used for clustering, it does not represent an independent validation of the clustering structure. In particular, the F statistic (ratio between-cluster variance and within-cluster variance) and its related p-value (probability, under the null hypothesis of equal means, of observing an F value at least this large) were calculated:

- White Blood Cells with $F = 257.35$ and $p = 8.55 \times 10^{-93}$, indicating significant differences in WBC counts across clusters;
- Age with $F = 151.17$ and $p = 6.88 \times 10^{-69}$, showing that age distributions differ significantly among the identified groups;
- Total IgE with $F = 186.33$ and $p = 7.06 \times 10^{-78}$, confirming a significant variation of IgE levels between clusters.

Significant differences were found for WBC, age and total IgE across clusters. This confirms that the consensus groups were statistically distinct with respect to all three variables used for clustering.

These ANOVA results should be interpreted as descriptive evidence of between-cluster differentiation within this cohort, rather than as an independent external validation of the clustering solution. Given the exploratory, post-clustering purpose of this analysis and the absence of external labels, ANOVA was used as a descriptive comparative tool rather than as an independent validation procedure.

Furthermore, all statistical analyses and clinical summaries were computed only on the subset of stable patients (287), while unassigned cases (13) were excluded.

4.4.3. Findings and Implications

The exploratory evaluation of the data showed how different variables interacted with each other. The middle ranges on the correlation matrix suggested that multivariate patterns would yield improved patient subgrouping outcomes compared to single-variable trends. In the case of the cluster analysis, the clustered features exhibited slight distribution asymmetry, hence the use of non-linear structure clustering methods became necessary.

These exploratory findings are confirmed by the consensus clusters and their clinical implications.

The consensus clusters generated preliminary clinically relevant distinctions between groups. Cluster 0 patients were typically older with higher WBC counts and moderate total IgE levels. In contrast, Cluster 1 contained younger patients with high IgE levels. Similarly, Cluster 2 included predominantly older individuals with relatively low WBC but elevated IgE levels, suggesting an allergic-dominant profile without marked systemic inflammation. Meanwhile, Cluster 3 represented younger patients with intermediate WBC and IgE values, indicating a relatively balanced immunological profile. Finally, Cluster 4 had the lowest total IgE levels, which may indicate a potentially hypoallergenic or immunologically less reactive group.

The ANOVA analysis indicated statistically significant between-cluster differences in WBC counts, age, and total IgE levels ($p < 0.0001$), supporting the descriptive differentiation of the identified profiles within this cohort. Additionally, the elbow method was used to determine the optimal number of clusters (k = 5), as illustrated in Figure 7.

The three-dimensional visualization of the data also showed an interpretable separation of the groups, which was deemed clinically appropriate by the allergists involved in the qualitative assessment of the results.

Indeed, the choice of k = 5 is consistent with previous phenotyping studies in allergic and asthmatic populations, which reported a comparable number of clinically significant

subgroups (typically four to six) [56–60]. In light of the above, our stratification into 5 clinical profiles is common and clinically sensible in the allergy field.

Taken together, these clusters suggest that similar biomarker values may correspond to different outpatient priorities depending on their combination with age and inflammatory context.

These profiles and their associated recommendations are designed for use in outpatient settings in Apulia, with adoption in other regions requiring consideration of local prevalence and clinical practice trends.

An unsupervised machine learning methodology enabled researchers to interpret clinical subgroups contained in actual patient database information. Robustness of the analysis resulted from using four clustering methods including K-means, Gaussian Mixture Models, Spectral Clustering and Hierarchical Clustering. Consensus clustering demanded the agreement of at least three algorithms to reduce algorithm-related biases and establish more constant and dependable patient clustering groups.

4.4.4. Clinical Interpretation of Biomarker-Based Clusters

From a clinical perspective, the identified clusters highlight the importance of interpreting immunological biomarkers within a broader biological and demographic context. Similar total IgE values may correspond to markedly different clinical risks depending on patient age and inflammatory status. For instance, elevated IgE levels in younger individuals may indicate an active and evolving allergic phenotype, whereas comparable IgE concentrations in older patients may reflect long-standing sensitization with limited current clinical impact. Conversely, increased white blood cell counts in older patients, even in the absence of markedly elevated IgE, may signal underlying inflammatory or comorbid processes that warrant further evaluation.

More specifically, the consensus clusters may reflect distinct clinical-biological profiles rather than mere numerical groupings. Younger patients with high total IgE and elevated WBC may be compatible with a more active immuno-allergic phenotype, whereas older patients with elevated WBC but only moderate IgE may suggest the coexistence of broader inflammatory or comorbidity-related processes beyond classical IgE-mediated mechanisms. Conversely, profiles with high IgE and relatively lower WBC may indicate predominant allergic sensitization without marked systemic inflammation, while clusters with lower values of both biomarkers may identify more stable, lower-priority outpatient profiles. Overall, these findings support the potential clinical relevance of the clusters for guiding follow-up intensity, test repetition, and referral prioritization. These groups should not be interpreted as diagnostic categories, but as clinically plausible phenotypic patterns derived from routine data.

4.5. Simplified Clinical Decisional Guide

A decision-support guide based on cluster characteristics is presented in Table 5. The table reports clinical reference ranges derived from consensus clusters and based only on patients assigned to stable clusters. In particular, this guide is tailored for outpatient workflows in Southern Italy and formatted for possible future integration into regional EMR systems.

The decision guidance was constructed drawing on evidence reported in the scientific literature for each cluster and subsequently reviewed for clinical terms, to ensure that the profiles and management recommendations associated with Clusters 0 [61–64], Clusters 1 [61,64–67], Clusters 2 [61,64,66], Clusters 3 [61–63,68,69] and Clusters 4 [61,64,68,70,71] reflect both consistency with age, WBC and total IgE data and practical relevance for patient management.

Table 5. Cluster profiles based on White Blood Cell count, total Immunoglobulin E, and Age interquartile ranges, ordered by increasing clinical priority, with corresponding clinical descriptions and outpatient management recommendations.

Cluster	Age Range (Years)	WBC (Cells/μL)	Total IgE (kU/L)	Clinical Profile	Recommended Clinical Management
4	55–61	6345–8091	137–172	Low systemic immune activity: • Older individuals with low immuno-allergic activity (low WBC and Total IgE); • No evidence of marked systemic inflammatory signs (based on total WBC count alone). This profile is consistent with a lower-priority outpatient phenotype.	Low priority: • Annual follow-up sufficient; • Low clinical priority unless other risks are identified.
3	39–46	7899–9769	232–272	Intermediate / stable profile: • Moderately aged patients with an overall balanced immunological profile (balanced WBC and Total IgE); • Clinically stable. This pattern may be compatible with an intermediate phenotype without clear predominance of allergic or inflammatory burden.	Moderate priority: • Semiannual review; • No immediate intervention unless symptoms evolve.
2	55–66	6392–7978	266–376	Predominantly allergic activity without marked systemic inflammation: • Mid-aged group with lower WBC but elevated Total IgE; • Atypical immune profile. This profile may identify patients needing mainly allergy-focused follow-up.	Allergy-focused priority: • Yearly check-up focusing on allergic responses and IgE levels; • Consider targeted diagnostic investigations as appropriate.
0	58–72	11,625–12,760	204–245	Possible inflammatory burden: • Older patients with persistent leukocytosis and moderate Total IgE levels, suggesting that inflammatory mechanisms may not be exclusively IgE-mediated; • This profile may be potentially indicative of non–IgE-driven inflammatory processes, including chronic low-grade inflammation and/or associated comorbidities, possibly age-related, and in some cases the contribution of latent inflammatory conditions, rather than isolated allergic activity. Biologically, this cluster may identify patients whose inflammatory burden is only partly explained by allergy.	Inflammatory/comorbidity-focused priority: • Biannual monitoring; • Check for chronic conditions, latent infections, or other sources of systemic inflammation.
1	35–42	12,017–12,888	310–335	High immuno-allergic activity: • Young patients with pronounced immuno-allergic activation (characterized by elevated Total IgE and leukocytosis) compatible with a possible immuno-allergic or immune-driven phenotype; • This pattern may be associated with a higher risk of disease progression, multimorbidity, or the development of new allergic manifestations over time. This pattern is compatible with a more active immuno-allergic phenotype and greater monitoring needs.	High immuno-allergic priority: • Regular allergy follow-up (at least every six months) with Total IgE re-evaluation; • Monitor for the onset of new symptoms or clinical deterioration.

Note: The profiles reported in this Table are intended as a pragmatic clinical risk stratification aid for outpatient management and follow-up prioritization. They should not be interpreted as diagnostic categories, but rather as biologically and clinically coherent patterns derived from routine laboratory and demographic data.

The table shows, for each stable cluster, the interquartile ranges (Q1–Q3) of Age, WBC, and total IgE: that is, the interval between the 25th and 75th percentiles, which describes the central 50% of patients in the cluster. These values are typical and robust ranges useful as a practical guide for recognizing similar clinical profiles.

Table 5 was developed and clinically reviewed by practicing allergists, who translated the cluster-wise biomarker ranges into pragmatic outpatient recommendations (based on the cited literature and their consolidated clinical experience).

Proposed follow-up intervals are based on the expected biological variability of total IgE and white blood cell counts, as well as on the potential clinical risk associated with each profile, rather than on rigid diagnostic thresholds.

This table is not intended to replace established clinical guidelines. Rather, it should be interpreted as a preliminary, decision aid that requires prospective validation before any routine clinical adoption.

4.6. Validation-Oriented Analysis: Supervised Proxy Prediction of Cluster Labels

This subsection provides a supervised proxy test of internal separability: if a simple classifier can accurately predict the unsupervised labels from the original features (Age, WBC and total IgE), this indicates that the partitions are coherent and well separated in feature space.

After unsupervised clustering (and label alignment via the Hungarian algorithm), the cluster assignments were treated as pseudo-labels. Accordingly, because the target labels were pseudo-labels derived from the same dataset, this supervised analysis should be interpreted only as an internal proxy assessment of cluster separability, rather than as a form of external validation. A supervised model was then trained to predict the cluster membership (i.e., the assignment to one of the five clusters) using the three features: Age, WBC, and total IgE.

In fact, a supervised learning model was trained using Random Forest Classifier to predict cluster labels. The dataset was split into training and testing sets with a 70:30 ratio. The Random Forest Classifier achieved the following result:

- Accuracy: 97.78%;
- Cross-Validation Accuracy Scores: [91.67%, 98.33%, 96.67%, 98.33%, 91.67%];
- Mean Cross-Validation Accuracy: 95.33%.

These results suggest that the cluster assignments are internally coherent and predictable within the analyzed dataset.

To further assess this internal separability, Stratified 5-Fold Cross-Validation was also considered. Under this setting, the Random Forest maintained high and stable performance (mean accuracy = 94.00% $\pm$ 4.42%; mean Macro F1 = 93.49% $\pm$ 4.78%), further supporting the internal separability of the pseudo-label-based clusters. The out-of-bag score was 0.919 (OOB error = 0.081), providing an additional internal estimate of predictive consistency.

It should be emphasized that this supervised analysis is not an external validation. Since the Random Forest was trained and tested on the same cohort used to derive the cluster labels, its performance mainly reflects internal separability of the identified groups rather than generalizability or direct clinical validity. External validation on independent cohorts is therefore still required.

Model Validation and Clinical Utility

The Random Forest model served for cluster validation but it underwent training and evaluation through an internal dataset split of 70/30 proportions. The model achieved 97.78% accuracy which demonstrates its capacity to replicate clustering patterns instead of showing generalization abilities for unknown clinical data.

High scores indicate that cluster boundaries are learnable by a simple model and therefore may reflect non-random structure in the analyzed feature space.

In particular, the classifier achieved high accuracy on the held-out split with misclassifications concentrated between clusters with adjacent ranges. This pattern supports the internal coherence of the discovered partitions.

The methodology based on Random Forest (RF) modeling for cluster validation does not serve as external verification since it operates within the same dataset and its 97.78 percent accuracy cannot prove clinical robustness. Future research should validate these clusters through testing with new patient populations because it is needed to prove their consistency across different healthcare groups.

In fact, given the regional nature of our cohort (Apulia region), external validation is important for assessing transportability to: (i) central and northern Italian regions; (ii) other health systems.

The researchers obtained profiles representing specific clinical characteristics together with forecasted comorbidities and proposed treatment plans for each cluster grouping. Healthcare professionals need these interpretations to convert machine learning outputs from abstraction into medical planning information. Some clusters consisted of individuals displaying characteristic high total IgE levels and young age but other clusters contained patients with older age and lower inflammatory markers. Healthcare professionals formulated different patient management approaches using these identification standards which included periodic follow-up timelines, evaluation systems and necessary referral criteria.

Healthcare providers can apply these practical guides to perform easy interpretation-based profiling of new patients through a simplified classification process that can avoid complex computational requirements.

Furthermore, because the pseudo-labels are derived from the same cohort, this test does not establish external validity or transportability; it only assesses internal separability. To mitigate optimism bias, one may re-fit clustering on the training portion only and then evaluate supervised prediction on the held-out set. External validation on independent cohorts remains necessary for clinical generalization.

A positive proxy test justifies deriving lightweight rules for EMR integration, while external datasets will be used to confirm transportability across regions.

5. Conclusions and Clinical Implications

This study identified patient groups based on WBC, age, and total IgE. Many different clustering algorithms were analyzed. Random Forest indicated internal separability of the identified groups through high classification accuracy.

5.1. Novelty from Clusters to Care

These conclusions were supported by an integrated workflow including data preparation, exploratory analysis, clustering and guided evaluation.

The study involved an in-depth clustering analysis to identify clinically relevant patient groups using three variables. There were several unsupervised models to secure credibility of different patterns that were identified.

The consensus clustering algorithm identified five distinct patient clusters that exhibited variably different profiles of WBC, age and total IgE level on the basis of the ANOVA test outcomes. The choice of k = 5 was guided by internal validation metrics, the elbow method, and clinical interpretability of the resulting profiles. Alternative values of k produced either overly aggregated groups or clinically less interpretable partitions.

An internal proxy assessment of cluster separability was performed with a Random Forest classifier, but this did not constitute clinical or external validation. Clinical interpretation linked the data-driven findings to practical outpatient management pathways.

Within the Southern Italy context, the proposed clustering framework may offer a low-cost and pragmatic support for preliminary outpatient stratification in routine primary care in Apulia.

Accordingly, a key element of this framework is its intentional parsimony: by relying on only three routine variables, the model prioritizes feasibility, affordability, and immediate clinical usability over exhaustive biological characterization. This design choice supports scalability and favors future adoption in resource-conscious outpatient settings.

In practical terms, the clustering results were used to derive outpatient recommendations for surveillance, diagnostic workup, and referral to specialists.

The following summary highlights the key points:

1. Minimal-cost outpatient risk stratification: using just three routine measures (age, white blood cell count, and total IgE), it is possible to stratify patients, distinguishing those requiring more frequent checkups from those who can be followed with scheduled surveillance (before resorting to specialist testing), with direct applicability in Apulia region.
   Specifically, five clinically interpretable strata were obtained via an ensemble-consensus rule (3/4 algorithms with label alignment by Hungarian matching), supported by internal indices (Silhouette, Calinski–Harabasz, Davies–Bouldin) and ANOVA on age, WBC, and total IgE ($p < 0.001$), potentially facilitating future EMR integration through simple reference ranges.
2. Reduction of variability between physicians: the proposed reference table standardizes operating thresholds and consequent actions, promoting more consistent decisions regarding when to reassess, intensify treatment, and refer for specialist consultation.
3. Translation into clinical actions: the derived profiles are operationalized into concrete outpatient steps:
   - Care pathways by phenotype: each profile identified by clustering is translated into practical recommendations: visit frequency, timing for repeat total IgE and white blood cell counts, opportunities for targeted in-depth testing (e.g., eosinophils, exhaled nitric oxide, spirometry), and criteria for referral to the most appropriate specialist. These pathways are intended for deployment in Southern Italian clinics;
   - Signs of subclinical inflammation in the elderly: the profile characterized by older age, elevated white blood cells, and only moderate total IgE suggests the possibility of non-clinically evident inflammation. In these cases, targeted screening (e.g., C-reactive protein, differential, airway evaluation, or chronic infections) is indicated to reduce potential diagnostic delays;
   - Prioritization of referrals: profiles that combine elevated white blood cell and IgE values at a younger age are associated with greater allergic activity and deserve priority for evaluation and treatment planning. Conversely, profiles with lower values can be managed with structured surveillance and periodic reassessment;
   - Monitoring over time: the evolution of an individual patient from one profile to another can be used as a pragmatic indicator of clinical response, useful both in daily practice.

5.2. Methodological Limitations and Future Directions

A key limitation is the use of a single-region primary-care cohort from Apulia; this may limit transportability to different healthcare systems.

Another limitation is the moderate sample size, which may affect cluster robustness and limit generalizability. The present findings should therefore be considered preliminary and validated in larger independent cohorts. Accordingly, the identified clusters should be interpreted as a preliminary data-driven subdivision of allergic patients, useful for generating clinically testable hypotheses and guiding subsequent clinical interpretation, rather than as definitive phenotypic categories. In addition, the Random Forest analysis mainly assesses internal separability, since cluster labels were derived from the same dataset used for classification. Therefore, it does not represent an independent validation of the stratification framework, and external validation on independent cohorts is still needed.

Accordingly, research should use external datasets accompanied by prospective studies to achieve complete validation of identified profiles regarding their value in personalized medicine and care systems.

To address this limitation, future external validation will require larger independent cohorts, clinically defined reference profiles, and the integration of additional biological markers, such as wheal size, specific IgE, and other immunological parameters.

Future work will also include a more detailed clinical analysis of the identified subgroups, using the present stratification as an initial framework to further assess their medical plausibility, associated comorbidities, and potential implications for patient management.

The proposed stratification model aims to provide a tool to support clinical reasoning and hypothesis generation, facilitating more consistent decisions in the outpatient setting, but it is not intended to replace established diagnostic and therapeutic guidelines in allergy.

Future work will expand the dataset to multi-regional Italian cohorts and international settings, enabling assessment of model transportability and fairness across diverse populations. A possible future direction is the adoption of fuzzy clustering approaches, which may better capture partially overlapping clinical profiles and graded membership patterns, whereas the present study was intentionally focused on obtaining clearly separated patient groups for interpretable clinical stratification.

Finally, more investigation should be conducted to understand the powerful link between patient age and immunoglobulin E levels from both physiological and immunological standpoints.

Author Contributions: Conceptualization, S.P., R.C. and C.L.; methodology, S.P., R.C. and C.L.; software, S.P.; validation, S.P., G.Z., R.C., C.L. and E.H.; formal analysis, S.P.; investigation, S.P. and G.Z.; clinical interpretation, E.H., A.U.G., G.Z. and S.P.; resources, G.Z.; data curation, G.Z. and S.P.; writing—original draft preparation, S.P.; writing—review and editing, S.P., E.H., A.U.G., G.Z., R.C. and C.L.; visualization, S.P.; supervision, E.H., R.C. and C.L.; project administration, C.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding

Institutional Review Board Statement: Ethical review and approval were waived for this study because it was a retrospective, observational, non-interventional analysis based exclusively on routinely collected clinical data. All data were irreversibly anonymized prior to analysis and no diagnostic or therapeutic procedures were modified for research purposes. In accordance with applicable Italian regulations and the Declaration of Helsinki, this type of study does not require formal Ethics Committee approval. The Ethics Committee of Universitas Mercatorum (Rome, Italy) issued a determination of no grounds to proceed and did not undertake a formal ethical evaluation.

Informed Consent Statement: Patient consent was waived due to the retrospective and non-interventional nature of the study and because all data were irreversibly anonymized prior to analysis, in accordance with GDPR (EU Regulation 2016/679) and applicable national regulations.

Data Availability Statement: In the interest of transparency and reproducibility, the anonymized dataset used in this study, provided as a de-identified Excel file including the variables relevant to the analyses, is available from the corresponding author upon request. The Python code developed for data preprocessing, clustering, validation, and the related analytical workflow is likewise available from the corresponding author upon request.

Acknowledgments: The authors acknowledge all those who provided support to this study.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
ANOVA	Analysis of Variance
ARI	Adjusted Rand Index
CH	Calinski–Harabasz Index
DB	Davies–Bouldin Index
EDA	Exploratory Data Analysis
EMR	Electronic Medical Record
EU	European Union
FMI	Fowlkes–Mallows Index
GDPR	General Data Protection Regulation
GMM	Gaussian Mixture Model
IgE	Immunoglobulin E
k-NN	k-Nearest Neighbors
NMI	Normalized Mutual Information
RF	Random Forest
SD	Standard Deviation
SVM	Support Vector Machine
WBC	White Blood Cell
WCSS	Within-Cluster Sum of Squares

References

1. Pawankar, R. Allergic diseases and asthma: A global public health concern and a call to action. *World Allergy Organ. J.* **2014**, *7*, 12. [CrossRef]
2. Calderón, M.A.; Linneberg, A.; Kleine-Tebbe, J.; De Blay, F.; Hernandez Fernandez de Rojas, D.; Virchow, J.C.; Demoly, P. Respiratory allergy caused by house dust mites: What do we really know? *J. Allergy Clin. Immunol.* **2015**, *136*, 38–48. [CrossRef]
3. Zhang, X.; Xing, F.; Zhao, Y.; Li, C. Efficacy of probiotics in the treatment of allergic diseases: A meta-analysis. *Front. Nutr.* **2025**, *12*, 1502390. [CrossRef]
4. Juniper, E.F.; Guyatt, G.H.; Dolovich, J. Assessment of quality of life in adolescents with allergic rhinoconjunctivitis: Development and testing of a questionnaire for clinical trials. *J. Allergy Clin. Immunol.* **1994**, *93*, 413–423. [CrossRef] [PubMed]
5. Baiardini, I.; Braido, F.; Bonini, M.; Compalati, E.; Canonica, G.W. Why do doctors and patients not follow guidelines? *Curr. Opin. Allergy Clin. Immunol.* **2009**, *9*, 228–233. [CrossRef] [PubMed]
6. Protudjer, J.L.P.; Davis, C.M.; Gupta, R.S.; Perry, T.T. Social determinants and quality of life in food allergy management and treatment. *J. Allergy Clin. Immunol. Pract.* **2025**, *13*, 745–750. [CrossRef] [PubMed]
7. Warren, C.M.; Otto, A.K.; Walkner, M.M.; Gupta, R.S. Quality of life among food allergic patients and their caregivers. *Curr. Allergy Asthma Rep.* **2016**, *16*, 38. [CrossRef]
8. Bousquet, J.; Mantzouranis, E.; Cruz, A.A.; Aït-Khaled, N.; Baena-Cagnani, C.E.; Bleecker, E.R.; Brightling, C.E.; Burney, P.; Bush, A.; Busse, W.W.; et al. Uniform definition of asthma severity, control, and exacerbations: Document presented for the World Health Organization Consultation on Severe Asthma. *J. Allergy Clin. Immunol.* **2010**, *126*, 926–938. [CrossRef]
9. Eguiluz-Gracia, I.; Tay, T.R.; Hew, M.; Escribese, M.M.; Barber, D.; O'Hehir, R.E.; Torres, M.J. Recent developments and highlights in biomarkers in allergic diseases and asthma. *Allergy* **2018**, *73*, 2290–2305. [CrossRef]

10. Ansotegui, I.J.; Melioli, G.; Canonica, G.W.; Caraballo, L.; Villa, E.; Ebisawa, M.; Passalacqua, G.; Savi, E.; Ebo, D.; Gómez, R.M.; et al. IgE allergy diagnostics and other relevant tests in allergy, a World Allergy Organization position paper. *World Allergy Organ. J.* **2020**, *13*, 100080. [CrossRef]
11. Vitte, J.; Santos, A.F. Editorial: In vitro diagnosis of allergic and mast cell-mediated disorders. *Front. Allergy* **2024**, *5*, 1483398 . [CrossRef]
12. Fritzsching, B. Personalized Medicine in Allergic Asthma: At the Crossroads of Allergen Immunotherapy and "Biologicals". *Front. Pediatr.* **2017**, *5*, 31. [CrossRef]
13. Muraro, A.; Fernandez-Rivas, M.; Beyer, K.; Cardona, V.; Clark, A.; Eller, E.; Hourihane, J.O.B.; Jutel, M.; Sheikh, A.; Agache, I.; et al. The urgent need for a harmonized severity scoring system for acute allergic reactions. *Allergy* **2018**, *73*, 1792–1800. [CrossRef] [PubMed]
14. Caddick, Z.A.; Fraundorf, S.H.; Rottman, B.M.; Nokes-Malach, T.J. Cognitive perspectives on maintaining physicians' medical expertise: II. Acquiring, maintaining, and updating cognitive skills. *Cogn. Res. Princ. Implic.* **2023**, *8*, 47. [CrossRef] [PubMed]
15. Niven, D.J.; McCormick, T.J.; Straus, S.E.; Hemmelgarn, B.R.; Jeffs, L.; Barnes, T.R.M.; Stelfox, H.T. Reproducibility of clinical research in critical care: A scoping review. *BMC Med.* **2018**, *16*, 26. [CrossRef] [PubMed]
16. Yang, W.C.; Lai, J.P.; Liu, Y.H.; Lin, Y.L.; Hou, H.P.; Pai, P.F. Using medical data and clustering techniques for a smart healthcare system. *Electronics* **2023**, *13*, 140. [CrossRef]
17. Aljuhani, M.; Ashraf, A.; Edison, P. Use of artificial intelligence in imaging dementia. *Cells* **2024**, *13*, 1965. [CrossRef]
18. Subramanian, J.; Simon, R. Gene expression-based prognostic signatures in lung cancer: Ready for clinical use? *J. Natl. Cancer Inst.* **2010**, *102*, 464–474. [CrossRef]
19. van Breugel, M.; Fehrmann, R.S.N.; Bügel, M.; Rezwan, F.I.; Holloway, J.W.; Nawijn, M.C.; Fontanella, S.; Custovic, A.; Koppelman, G.H. Current state and prospects of artificial intelligence in allergy. *Allergy* **2023**, *78*, 2623–2643. [CrossRef]
20. Haider, S.; Fontanella, S.; Ullah, A.; Turner, S.; Simpson, A.; Roberts, G.; Murray, C.S.; Holloway, J.W.; Curtin, J.A.; Cullinan, P.; et al. Evolution of eczema, wheeze, and rhinitis from infancy to early adulthood: Four birth cohort studies. *Am. J. Respir. Crit. Care Med.* **2022**, *206*, 950–960. [CrossRef]
21. Monti, S.; Tamayo, P.; Mesirov, J.; Golub, T. Consensus Clustering: A Resampling-Based Method for Class Discovery and Visualization of Gene Expression Microarray Data. *Mach. Learn.* **2003**, *52*, 91–118. [CrossRef]
22. Kiselev, V.Y.; Kirschner, K.; Schaub, M.T.; Andrews, T.; Yiu, A.; Chandra, T.; Natarajan, K.N.; Reik, W.; Barahona, M.; Green, A.R.; et al. SC3: Consensus clustering of single-cell RNA-seq data. *Nat. Methods* **2017**, *14*, 483–486. [CrossRef]
23. Fred, A.L.; Jain, A.K. Combining multiple clusterings using evidence accumulation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 835–850. [CrossRef]
24. Ahmed, M.B.; Ad'hiah, A.H. Effects of age, gender and allergen type on immunoglobulin E level in asthma and allergic rhinitis patients. *Iraqi J. Sci.* **2022**, *63*, 1498–1506. [CrossRef]
25. Al-Ghamdi, B.R.; Koshak, E.A.; Omer, F.M.; Awadalla, N.J.; Mahfouz, A.A.; Ageely, H.M. Immunological factors associated with adult asthma in the Aseer region, southwestern Saudi Arabia. *Int. J. Environ. Res. Public Health* **2019**, *16*, 2495. [CrossRef]
26. Leung, T.F.; Kong, A.P.S.; Chan, I.H.S.; Choi, K.C.; Ho, C.S.; Chan, M.H.M.; So, W.Y.; Lam, C.W.K.; Wong, G.W.K.; Chan, J.C.N. Association between obesity and atopy in Chinese schoolchildren. *Int. Arch. Allergy Immunol.* **2009**, *149*, 133–140. [CrossRef]
27. Guo, J.; Huang, X.; Dou, L.; Yan, M.; Shen, T.; Tang, W.; Li, J. Aging and aging-related diseases: From molecular mechanisms to interventions and treatments. *Signal Transduct. Target. Ther.* **2022**, *7*, 391. [CrossRef]
28. Chmielewski, P.P.; Strzelec, B. Elevated leukocyte count as a harbinger of systemic inflammation, disease progression, and poor prognosis: A review. *Folia Morphol.* **2018**, *77*, 171–178. [CrossRef]
29. Amarasekera, M. Immunoglobulin E in health and disease. *Asia Pac. Allergy* **2011**, *1*, 12–15. [CrossRef] [PubMed]
30. Junod, V.; Elger, B. Retrospective research: What are the ethical and legal requirements? *Swiss Med. Wkly.* **2010**, *140*, w13041. [CrossRef] [PubMed]
31. Winter, E.M.; Maughan, R.J. Requirements for ethics approvals. *J. Sport. Sci.* **2009**, *27*, 985. [CrossRef] [PubMed]
32. Gollogly, L. Ethical approval for operational research. *Bull. World Health Organ.* **2006**, *84*, 766 . [CrossRef]
33. Wani, A.A. Comprehensive analysis of clustering algorithms: Exploring limitations and innovative solutions. *PeerJ Comput. Sci.* **2024**, *10*, e2286. [CrossRef] [PubMed]
34. Huang, H.; Liao, Z.; Wei, X.; Zhou, Y. Combined Gaussian mixture model and Pathfinder algorithm for data clustering. *Entropy* **2023**, *25*, 946. [CrossRef] [PubMed]
35. Bertsimas, D.; Orfanoudaki, A.; Wiberg, H. Interpretable clustering: An optimization approach. *Mach. Learn.* **2021**, *110*, 89–138. [CrossRef]
36. Javed, H.; El-Sappagh, S.; Abuhmed, T. Robustness in deep learning models for medical diagnostics: Security and adversarial challenges towards robust AI applications. *Artif. Intell. Rev.* **2024**, *58*, 12. [CrossRef]
37. Boutalbi, R.; Labiod, L.; Nadif, M. Implicit consensus clustering from multiple graphs. *Data Min. Knowl. Discov.* **2021**, *35*, 2313–2340. [CrossRef]

38. Tangherloni, A.; Ricciuti, F.; Besozzi, D.; Liò, P.; Cvejic, A. Analysis of single-cell RNA sequencing data based on autoencoders. *BMC Bioinform.* **2021**, *22*, 309. [CrossRef]
39. Handschuh, L.; Kaźmierczak, M.; Milewski, M.; Góralski, M.; Łuczak, M.; Wojtaszewska, M.; Uszczyńska-Ratajczak, B.; Lewandowski, K.; Komarnicki, M.; Figlerowicz, M. Gene expression profiling of acute myeloid leukemia samples from adult patients with AML-M1 and -M2 through boutique microarrays, real-time PCR and droplet digital PCR. *Int. J. Oncol.* **2017**, 656–678. [CrossRef]
40. Karras, C.; Karras, A.; Giotopoulos, K.C.; Avlonitis, M.; Sioutas, S. Consensus big data clustering for Bayesian mixture models. *Algorithms* **2023**, *16*, 245. [CrossRef]
41. Gabrovšek, B.; Novak, T.; Povh, J.; Rupnik Poklukar, D.; Žerovnik, J. Multiple Hungarian method for k-assignment problem. *Mathematics* **2020**, *8*, 2050. [CrossRef]
42. Hu, J.; Szymczak, S. A review on longitudinal data analysis with random forest. *Brief. Bioinform.* **2023**, *24*, bbad002.
43. Sarica, A.; Cerasa, A.; Quattrone, A. Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: A systematic review. *Front. Aging Neurosci.* **2017**, *9*, 329. [CrossRef] [PubMed]
44. Musa, A.B. Comparative study on classification performance between support vector machine and logistic regression. *Int. J. Mach. Learn. Cybern.* **2013**, *4*, 13–24. [CrossRef]
45. Ben-Hur, A.; Weston, J. A User's Guide to Support Vector Machines. In *Methods in Molecular Biology*; Clifton, N.J., Ed.; Humana Press: Totowa, NJ, USA, 2010; pp. 223–239.
46. Gelman, A.; Jakulin, A.; Pittau, M.G.; Su, Y.S. A weakly informative default prior distribution for logistic and other regression models. *Ann. Appl. Stat.* **2008**, *2*, 1360–1383. [CrossRef]
47. Friedman, J.; Hastie, T.; Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **2010**, *33*, 1–22. [CrossRef]
48. Hu, L.Y.; Huang, M.W.; Ke, S.W.; Tsai, C.F. The distance function effect on k-nearest neighbor classification for medical datasets. *Springerplus* **2016**, *5*, 1304. [CrossRef] [PubMed]
49. Probst, P.; Wright, M.; Boulesteix, A.L. Hyperparameters and tuning strategies for random forest. *arXiv* **2018**, arXiv:1804.03515. [CrossRef]
50. Lange, T.M.; Gültas, M.; Schmitt, A.O.; Heinrich, F. optRF: Optimising random forest stability by determining the optimal number of trees. *BMC Bioinform.* **2025**, *26*, 95. [CrossRef] [PubMed]
51. Zhou, S.; Mentch, L. Trees, forests, chickens, and eggs: When and why to prune trees in a random forest. *arXiv* **2021**, arXiv:2103.16700. [CrossRef]
52. Nadi, A.; Moradi, H. Increasing the views and reducing the depth in random forest. *Expert Syst. Appl.* **2019**, *138*, 112801. [CrossRef]
53. Saito, T.; Rehmsmeier, M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS ONE* **2015**, *10*, e0118432. [CrossRef] [PubMed]
54. Nembrini, S.; König, I.R.; Wright, M.N. The revival of the Gini importance? *Bioinformatics* **2018**, *34*, 3711–3718. [CrossRef]
55. Dunne, R.; Reguant, R.; Ramarao-Milne, P.; Szul, P.; Sng, L.M.F.; Lundberg, M.; Twine, N.A.; Bauer, D.C. Thresholding Gini variable importance with a single-trained random forest: An empirical Bayes approach. *Comput. Struct. Biotechnol. J.* **2023**, *21*, 4354–4360. [CrossRef]
56. Lötvall, J.; Akdis, C.A.; Bacharier, L.B.; Bjermer, L.; Casale, T.B.; Custovic, A.; Lemanske, R.F., Jr.; Wardlaw, A.J.; Wenzel, S.E.; Greenberger, P.A. Asthma endotypes: A new approach to classification of disease entities within the asthma syndrome. *J. Allergy Clin. Immunol.* **2011**, *127*, 355–360. [CrossRef]
57. Moore, W.C.; Meyers, D.A.; Wenzel, S.E.; Teague, W.G.; Li, H.; Li, X.; D'Agostino, R., Jr.; Castro, M.; Curran-Everett, D.; Fitzpatrick, A.M.; et al. Identification of asthma phenotypes using cluster analysis in the Severe Asthma Research Program. *Am. J. Respir. Crit. Care Med.* **2010**, *181*, 315–323. [CrossRef]
58. Haldar, P.; Pavord, I.D.; Shaw, D.E.; Berry, M.A.; Thomas, M.; Brightling, C.E.; Wardlaw, A.J.; Green, R.H. Cluster analysis and clinical asthma phenotypes. *Am. J. Respir. Crit. Care Med.* **2008**, *178*, 218–224. [CrossRef] [PubMed]
59. Loza, M.; Adcock, I.; Auffray, C.; Chung, K.F.; Djukanovic, R.; Sterk, P.; Susulic, V.; Barnathan, E.; Baribaud, F.; Silkoff, P. Longitudinally Stable, Clinically Defined Clusters of Patients with Asthma Independently Identified in the ADEPT and U-BIOPRED Asthma Studies. *Ann. Am. Thorac. Soc.* **2016**, *13*, S102–S103. [CrossRef] [PubMed]
60. Bousquet, J.; Anto, J.; Auffray, C.; Akdis, M.; Cambon-Thomsen, A.; Keil, T.; Haahtela, T.; Lambrecht, B.; Postma, D.; Sunyer, J.; et al. MeDALL (Mechanisms of the Development of ALLergy): An integrated approach from phenotypes to systems medicine. *Allergy* **2011**, *66*, 596–604. [CrossRef]
61. Guida, G.; Bertolini, F.; Carriero, V.; Levra, S.; Sprio, A.E.; Sciolla, M.; Orpheu, G.; Arrigo, E.; Pizzimenti, S.; Ciprandi, G.; et al. Reliability of total serum IgE levels to define type 2 high and low asthma phenotypes. *J. Clin. Med.* **2023**, *12*, 5447. [CrossRef]
62. Chuang, Y.C.; Tsai, H.H.; Lin, M.C.; Wu, C.C.; Lin, Y.C.; Wang, T.N. Cluster analysis of phenotypes, job exposure, and inflammatory patterns in elderly and nonelderly asthma patients. *Allergol. Int.* **2024**, *73*, 214–223. [CrossRef]

63. Ilmarinen, P.; Tuomisto, L.E.; Niemelä, O.; Tommola, M.; Haanpää, J.; Kankaanranta, H. Cluster analysis on longitudinal data of patients with adult-onset asthma. *J. Allergy Clin. Immunol. Pract.* **2017**, *5*, 967–978.e3. [CrossRef]
64. Carr, T.F.; Kraft, M. Use of biomarkers to identify phenotypes and endotypes of severeasthma. *Ann. Allergy Asthma Immunol.* **2018**, *121*, 414–420. [CrossRef]
65. Wu, W.; Bleecker, E.; Moore, W.; Busse, W.W.; Castro, M.; Chung, K.F.; Calhoun, W.J.; Erzurum, S.; Gaston, B.; Israel, E.; et al. Unsupervised phenotyping of Severe Asthma Research Program participants using expanded lung data. *J. Allergy Clin. Immunol.* **2014**, *133*, 1280–1288. [CrossRef]
66. Sendín-Hernández, M.P.; Ávila-Zarza, C.; Sanz, C.; García-Sánchez, A.; Marcos-Vadillo, E.; Muñoz-Bellido, F.J.; Laffond, E.; Domingo, C.; Isidoro-García, M.; Dávila, I. Cluster analysis identifies 3 phenotypes within allergic asthma. *J. Allergy Clin. Immunol. Pract.* **2018**, *6*, 955–961.e1. [CrossRef]
67. Lee, E.; Hong, S.J. Phenotypes of allergic diseases in children and their application in clinical situations. *Korean J. Pediatr.* **2019**, *62*, 325–333. [CrossRef]
68. Loza, M.J.; Djukanovic, R.; Chung, K.F.; Horowitz, D.; Ma, K.; Branigan, P.; Barnathan, E.S.; Susulic, V.S.; Silkoff, P.E.; Sterk, P.J.; et al. Validated and longitudinally stable asthma phenotypes based on cluster analysis of the ADEPT study. *Respir. Res.* **2016**, *17*, 165. [CrossRef] [PubMed]
69. Nadif, R.; Febrissy, M.; Andrianjafimasy, M.V.; Le Moual, N.; Gormand, F.; Just, J.; Pin, I.; Siroux, V.; Matran, R.; Dumas, O.; et al. Endotypes identified by cluster analysis in asthmatics and non-asthmatics and their clinical characteristics at follow-up: The case-control EGEA study. *BMJ Open Respir. Res.* **2020**, *7*, e000632. [CrossRef] [PubMed]
70. Lee, Y.; Quoc, Q.L.; Park, H.S. Biomarkers for severe asthma: Lessons from longitudinal cohort studies. *Allergy Asthma Immunol. Res.* **2021**, *13*, 375–389. [CrossRef] [PubMed]
71. Denton, E.; Price, D.B.; Tran, T.N.; Canonica, G.W.; Menzies-Gow, A.; FitzGerald, J.M.; Sadatsafavi, M.; Perez de Llano, L.; Christoff, G.; Quinton, A.; et al. Cluster analysis of inflammatory biomarker expression in the International Severe Asthma Registry. *J. Allergy Clin. Immunol. Pract.* **2021**, *9*, 2680–2688.e7. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.