

Clinical Phenotypes of Intracranial Suppurative Infections Identified by Unsupervised Machine Learning

Koffi Yves Soress Dongo^{1,2*}, Faozo Stéphane Landry Teti^{1,2}, Kouadio Jean-Baptiste Keke^{1,2},
Yao Bernard Fionko^{1,2}, Aderehime Haidara^{1,2}

¹Department of Neurosurgery, Bouaké University Hospital, Bouaké, Côte d'Ivoire

²Faculty of Medical Sciences, Alassane Ouattara University, Bouaké, Côte d'Ivoire

Email: *dongosoress0723@gmail.com

How to cite this paper: Dongo, K.Y.S., Teti, F.S.L., Keke, K.J.-B., Fionko, Y.B. and Haidara, A. (2026) Clinical Phenotypes of Intracranial Suppurative Infections Identified by Unsupervised Machine Learning. *Neuroscience and Medicine*, 17, 169-210. <https://doi.org/10.4236/nm.2026.174014>

Received: September 20, 2026

Accepted: October 11, 2026

Published: October 14, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Background: Intracranial suppurative infections (ISIs) are conventionally classified by anatomical compartment—brain abscess, subdural empyema, extradural empyema—a taxonomy that captures neither the heterogeneity of routes of infection nor the diversity of care pathways. No previous study has examined whether unsupervised learning can redefine clinically meaningful phenotypes of ISI beyond conventional anatomical categories, and none has done so in an African cohort. **Methods:** We analysed 150 consecutive observations (148 patients) of ISI managed at Bouaké University Hospital, Côte d'Ivoire, between 2016 and 2026. In the primary analysis, nineteen pre-therapeutic variables spanning host factors, symptom-to-admission delay, neurological severity, grouped portal of entry, lesion type and lesion burden were entered into a factor analysis of mixed data (FAMD); the origin of the suppuration, almost collinear with the portal of entry, was removed, and the specification including both was kept as a sensitivity analysis. Phenotypes were derived by K-means clustering on the retained factor coordinates, with Ward clustering on the same Euclidean coordinates and partitioning around medoids (PAM) on Gower distances as concordance checks. The reproducibility of the whole pipeline was assessed by repeated nested cross-validation, imputation, FAMD and K-means being re-estimated in every training set. In-hospital mortality and functional status were analysed as separate outcomes, and the incremental prognostic value of phenotype beyond GCS and age was assessed with optimism correction by bootstrap (1000 replicates). **Results:** In the primary non-redundant specification, a four-class solution was retained (silhouette 0.296). Its concordance with the dual-coded partition was moderate (adjusted Rand index [ARI] 0.325), and the frontal-interhemispheric sino-

genic subdural empyema phenotype was the most robust (28 of 33 observations recovered). PAM on Gower distances identified four groups of 48, 36, 34 and 32 observations (silhouette 0.312; ARI with FAMD-K-means 0.671). Out-of-sample agreement of the complete pipeline was $87.6\% \pm 4.8\%$ (mean ARI 0.79 ± 0.09). The dual-coded specification, used for detailed characterisation, separated community-acquired brain abscess of ENT-dental or cryptogenic origin ($n = 54$), delayed post-traumatic suppuration ($n = 33$; median delay 29 days), sinogenic subdural empyema ($n = 33$; craniotomy 91%, external ventricular drainage 24%) and post-operative suppuration ($n = 30$; Gram-negative-enriched microbiology). In-hospital mortality ranged from 5.6% to 18.2% across phenotypes ($p = 0.29$). Among 126 evaluable survivors, 41 (32.5%) had a GOS ≤ 3 , ranging from 19.2% to 40.0% across phenotypes ($p = 0.38$). Adding phenotype to GCS and age raised the optimism-corrected AUC by 0.009 only, for mortality (0.754 to 0.763) as for functional status (0.794 to 0.803), without gain in Brier score or calibration. **Conclusions:** Unsupervised learning delineated clinically recognisable ISI phenotypes, organised mainly by the mechanism of infection and the care pathway and corroborated by a microbiological distribution that was never used to build them. Once the dual coding of the mechanism was removed, only the sinogenic subdural empyema phenotype was consistently reproduced; the other groupings remained conditional on the weight given to the mechanism. Phenotype added little prognostic information to admission GCS and age. Phenotype-aware triage, rather than phenotype-based prognostication, is the defensible clinical application of these results, and external validation in multicentre African cohorts is required before implementation.

Keywords

Intracranial Suppuration, Brain Abscess, Subdural Empyema, Unsupervised Machine Learning, Clinical Phenotypes, Côte d'Ivoire

1. Introduction

Intracranial suppurative infections (ISIs) form a heterogeneous group of focal infections of the central nervous system, comprising brain abscess, subdural empyema, extradural empyema, ventriculitis and pre-suppurative cerebritis [1]–[3]. Despite advances in cross-sectional imaging and antibiotic therapy, their mortality remains between 10% and 20% in contemporary series, and neurological sequelae affect nearly one survivor in two [4]–[8]. The European guidelines published in 2024 clarified the diagnostic and therapeutic strategy for brain abscess [9], but rest almost exclusively on data from high-income countries.

Yet it is in sub-Saharan Africa that a major share of the burden of this condition is concentrated, under very different conditions of practice. The *Lancet Commission on Global Surgery* established that five billion people have no access to safe, affordable surgery [10], and the analysis of global neurosurgical capacity esti-

mated at more than five million the annual number of essential neurosurgical operations not performed, the great majority in Africa and South-East Asia [11]. Intracranial suppurative infections belong precisely to the core of essential neurosurgical conditions—curable at low cost and severely disabling in the absence of care—whose integration into universal health coverage benefit packages is now recommended [12]–[14]. In such settings, mortality and sequelae stem less from the sophistication of the procedure than from delays in seeking care, fragmentation of pathways and the scarcity of microbiological documentation [15]–[20].

The usual classification of ISIs is anatomical. It separates collections according to the compartment involved and, on that basis, guides the surgical procedure. It nonetheless accounts imperfectly for clinical reality: a sinogenic subdural empyema in a child, an otogenic temporal abscess in an adult and a surgical-site supuration after craniotomy may share a compartmental label, but they share neither the same portal of entry, nor the same kinetics, nor the same microbial ecology, nor the same care pathway [19] [21]–[27]. Conversely, lesions in different compartments may arise from a common causal mechanism. This mismatch between lesional taxonomy and pathophysiological reality limits the comparability of series and the design of trials.

The great majority of work devoted to ISIs uses supervised methods: an outcome variable is fixed a priori—most often mortality—and regression models seek its determinants. This approach has produced a robust, repeatedly confirmed result, the prognostic value of the initial Glasgow score [6] [7] [28], but it is structurally unable to reveal the internal organisation of the population studied: it describes the relationship between variables and an outcome, not the natural groupings of patients. Unsupervised learning answers a different and complementary question: with no outcome variable, it lets the multidimensional configuration of the observations itself determine the groupings [29]. This approach has profoundly renewed the description of sepsis [30], acute respiratory distress syndrome [31] [32], adult diabetes [33], asthma [34] and heart failure with preserved ejection fraction [35], by identifying subgroups whose mechanism, trajectory and sometimes treatment response differed to the point that some trials reanalysed a posteriori revealed opposite treatment effects according to subphenotype [32]. Its value in neurosurgical infectious disease has never been assessed.

To our knowledge, no published study has examined whether unsupervised learning could redefine clinically meaningful phenotypes of intracranial suppurative infections beyond conventional anatomical classifications, and no African cohort of ISI has been the subject of such an analysis. We use *clinical phenotype* here to denote a reproducible subgroup of patients sharing similar observable characteristics, independently of the underlying biological mechanisms [29]. In this work, phenotypes correspond to the latent classes identified by the unsupervised algorithm; the term is used in preference to class, group or *cluster*, the latter used only where the technical context requires it. Finally, it is important to distinguish prediction from phenotyping: the former estimates an outcome fixed a priori from

input variables, the latter describes the latent organisation of a population with no outcome variable. This work belongs exclusively to the second register.

The aim of this work was to identify, by unsupervised classification applied to admission data alone, the clinical phenotypes naturally present within a monocentric Ivorian cohort, to characterise their epidemiological, lesional, therapeutic, microbiological and outcome profile, and to assess whether this classification added prognostic information beyond the initial Glasgow score.

2. Patients and Methods

2.1. Study Type and Setting, Participants

This was an observational cohort study, both descriptive and analytical, with mixed retrospective and prospective recruitment, conducted over a ten-and-a-half-year period from 1 January 2016 to 30 June 2026 in the Department of Neurosurgery of Bouaké University Hospital (Côte d'Ivoire). The retrospective phase covered inclusions up to 31 December 2024; data collection became prospective from 1 January 2025. The analysis presented here is a secondary, exploratory analysis of the database assembled for this cohort. All patients managed for an intracranial suppurative infection documented clinically and by cross-sectional imaging (computed tomography and/or MRI) were included, whatever the age, the portal of entry and the therapeutic strategy. No eligible episode was excluded. The unit of analysis was the observation, *i.e.*, the hospitalisation episode; two patients having presented two distinct episodes, 150 observations corresponding to 148 patients were analysed.

The sample size was not determined a priori: it corresponds to the exhaustive set of eligible episodes recorded in the centre over a decade. This exhaustiveness limits the risk of selection within the hospital cohort, but excludes neither a referral bias, nor an access-to-care bias, nor the non-inclusion of patients who died before transfer or who never accessed imaging; the series therefore constitutes a hospital sample of the broader target population of intracranial suppurative infections occurring in the region.

There is no universally accepted method to determine a priori the sample size required for an unsupervised classification analysis: the stability of a partition depends on the number and type of variables, on the actual separability of the groups and on the algorithm used [36] [37], and the available formal rules concern supervised prognostic models [38]. Given the ratio of sample size to the number of candidate variables, a prior dimensionality reduction and several complementary robustness assessments were implemented.

The report follows the STROBE recommendations [39] [40]. For the machine-learning component, the transparency principles of the TRIPOD + AI framework relating to preprocessing, the choice of the number of classes and the assessment of stability were applied [41], and the relevant items of PROBAST + AI [42] and of the FUTURE-AI framework [43] guided the description of robustness and explainability. It should be stressed, however, that these three frameworks, like MI-

CLAIM [44], DECIDE-AI [45], CONSORT-AI [46] and SPIRIT-AI [47], were designed for supervised prognostic models and for the clinical evaluation of AI-based interventions; none applies literally to a descriptive unsupervised analysis. They were therefore used as a transparency framework and not as a compliance checklist, and no claim of formal conformity is made.

2.2. Variables Retained for the Classification

The guiding principle of the selection was priority over the therapeutic decision. Only variables available at the end of the initial work-up (history, admission clinical examination, first cross-sectional imaging) were retained, to the exclusion of any therapeutic, microbiological or outcome variable. This choice avoids the circularity of deriving classes from management or outcome and then comparing them on those same dimensions; it also reserves the microbiological data for the role of external validation.

Twenty candidate variables were considered, spanning five domains (**Table 1**): host factors (age, immunocompromise), the symptom-to-admission delay, initial neurological severity (GCS, impaired consciousness, focal motor deficit, seizures, fever), the portal of entry and origin of the suppuration, and lesion burden (type of collection, ventriculitis, scalp suppuration, maximum diameter, multiplicity, mass effect, hydrocephalus, topography). The origin of the suppuration and the grouped portal of entry proved almost deterministically associated (Cramér's $V = 0.989$), the latter being a refinement of the former. To avoid giving a double weight to the mechanism of infection, the primary analysis used nineteen variables: the origin was removed and the grouped portal of entry retained. The twenty-variable specification including both, hereafter called the dual-coded specification, was kept as a sensitivity analysis. Variables of extreme prevalence (<3% or >90%), redundant variables, or those with a missing-data rate above 20% were excluded from the classification phase and reserved for post-hoc characterisation. Portals of entry were grouped into five categories (ENT-dental, post-traumatic, post-operative, cutaneous/scalp, haematogenous/cryptogenic) and topographies into five categories, in order to avoid categories with zero counts within the classes.

2.3. Missing Data

Among the twenty classification variables, only one contained missing data: the maximum lesion diameter (6 values, *i.e.* 4.0% of observations and 0.2% of all cells). No other retained variable was incomplete.

These values were imputed by the cohort median. This choice, ordinarily debatable, is justified here by three arguments. First, multiple imputation by chained equations [48] or by random forests [49] derives its superiority from exploiting correlations between variables and from propagating imputation uncertainty into subsequent estimates [50]; yet the primary analysis is a classification, not an inferential procedure producing standard errors, so that the main benefit of multiple imputation—appropriately widened confidence intervals—is moot. Second,

Table 1. Candidate variables for the unsupervised classification (n = 20; 19 in the primary specification).

Domain	Variable	Type	Missing data
Host factors	Age (years)	Quantitative	0%
Host factors	Immunocompromise	Binary	0%
Delay	Symptom-to-admission delay (days)	Quantitative	0%
Clinical severity	Admission GCS	Quantitative	0%
Clinical severity	Impaired consciousness	Binary	0%
Clinical severity	Focal motor deficit	Binary	0%
Clinical severity	Seizures	Binary	0%
Clinical severity	Fever	Binary	0%
Origin	Origin of the suppuration, sensitivity analysis only (3 categories)	Categorical	0%
Origin	Grouped portal of entry (5 categories)	Categorical	0%
Lesional	Brain abscess	Binary	0%
Lesional	Subdural empyema	Binary	0%
Lesional	Extradural empyema	Binary	0%
Lesional	Ventriculitis	Binary	0%
Lesional	Scalp suppuration	Binary	0%
Lesional	Mass effect	Binary	0%
Lesional	Hydrocephalus	Binary	0%
Lesional	Multiple collections	Binary	0%
Lesional	Maximum lesion diameter (mm)	Quantitative	4.0%
Lesional	Grouped topography (5 categories)	Categorical	0%

Categories of the categorical variables—Origin: community/spontaneous, post-traumatic, post-operative. Portal of entry: ENT-dental, post-traumatic, post-operative, cutaneous/scalp, haematogenous/cryptogenic. Topography: frontal, parietal, temporal, interhemispheric, other/posterior.

multiple imputation would have produced *m* datasets potentially associated with different partitions; although consensus methods exist, their implementation is not standardised for this type of exploratory analysis and would have introduced complexity disproportionate to the observed missing-data rate. Third, and decisively, the proportion of affected values is too small for the choice of method to influence the structure: this point was verified empirically (§ 2.8).

2.4. Dimensionality Reduction

The mixed nature of the data—four quantitative variables, thirteen binary variables and two categorical variables in the primary specification, three in the dual-coded specification—precluded a classical principal component analysis. A factor

analysis of mixed data (FAMD) was therefore performed [51], which generalises PCA to quantitative variables and multiple correspondence analysis to qualitative variables, weighting each group so that none mechanically dominates the first axes. Quantitative variables were standardised, and qualitative categories were weighted by the inverse of their frequency.

FAMD was preferred to principal component analysis, inapplicable to qualitative variables, and to multiple correspondence analysis alone, which would have required arbitrarily discretising the quantitative variables and losing the ordering information; it was also preferred to multiple factor analysis, suited when variables are organised into a priori thematic groups whose influence one wishes to balance, which was not our setting. FAMD jointly models continuous and qualitative variables while balancing their respective contributions to the latent space, preventing one data type from mechanically dominating the solution.

The number of retained components was set at four, in the absence of a clear break in the scree plot, on the basis of the stability of the classification solutions obtained on 3 to 8 components. In the dual-coded specification, these four axes concentrated 37.9% of the total inertia.

2.5. Unsupervised Classification

The primary classification was performed by the K-means algorithm applied to the factor coordinates of the first four FAMD components, with 500 random initialisations (k-means++ method) and a fixed seed, to ensure reproducibility.

The K-means algorithm has known limitations: it assumes approximately spherical classes of comparable dispersion, it is sensitive to outliers and it requires the number of classes to be fixed a priori. It was nonetheless retained for its transparency, its reproducibility and its wide use in clinical phenotyping [29], and its limitations were offset by systematic comparison with methods not sharing the same assumptions: Ward hierarchical clustering, which imposes no shape on the classes, and a clustering on Gower distances computed without dimensionality reduction. Gaussian mixture models or a latent class analysis would have been legitimate alternatives; however, their implementation on 150 observations and twenty variables would have entailed a higher risk of over-parameterisation.

Two concordance analyses were performed. The first was a hierarchical agglomerative classification using the Ward criterion [52], applied to the same factor coordinates and assessed by the cophenetic correlation. The second was a partitioning around medoids (PAM) [53] on a matrix of Gower distances [54], computed directly on the nineteen raw variables of the primary specification, without dimensionality reduction. Ward's criterion minimises the within-cluster variance and therefore requires Euclidean distances; it was applied only to the FAMD factor coordinates and no longer to Gower dissimilarities. The PAM analysis constitutes the most demanding robustness test, since it borrows nothing from the FAMD. Concordance between partitions was quantified by the adjusted Rand index (ARI) [55].

2.6. Choice of the Number of Classes and Assessment of Stability

The number of classes was determined jointly by five criteria: the silhouette coefficient, mean and per class [56], the Calinski-Harabasz index [57], the Davies-Bouldin index [58], the gap statistic [59] with 100 uniform reference datasets, and the clinical interpretability of the solutions.

Stability was assessed class by class using Hennig's method [60], by non-parametric bootstrap resampling (500 resamples). For each resample, a new K-class classification was computed, then the whole set of original observations—including those not drawn—was re-assigned to the nearest centroid of this new solution, in accordance with the supervised classification step provided for by the procedure. The maximum Jaccard index between each reference class and the classes of the resample was then computed, then summarised by its mean and its distribution. By the usual thresholds, a value above 0.85 indicates a highly stable class, a value between 0.75 and 0.85 a stable class, between 0.60 and 0.75 a real but imperfectly delineated structure, and below 0.60 an unstable class.

The supervised separability of the phenotypes was examined descriptively by a random forest [61] of 2000 trees re-assigning each observation to its class from the base variables alone. Its performance was assessed by out-of-bag error and by ten-fold stratified cross-validation, with overall and balanced accuracy, area under the ROC curve in a one-versus-all approach, multiclass Brier score [62] and confusion matrix. As the phenotype labels had been determined once and for all on the complete cohort, cross-validation assessed only the ability of the supervised classifier to reproduce these fixed labels: this is not a nested cross-validation of the whole phenotyping pipeline. A high accuracy therefore indicates that the assignment obtained by the unsupervised procedure can be reliably approximated by a supervised function of the initial variables—*i.e.* that the classes are internally separable—but demonstrates neither their reproducibility in an independent population, nor their prospective clinical utility. Variable importance (mean decrease in impurity) is reported descriptively and must be interpreted with caution, this measure being sensitive to the type and number of categories of the variables.

Because such fixed labels already carry information from the observations held out in each fold, the internal reproducibility of the complete phenotyping pipeline was assessed separately by repeated nested cross-validation (ten folds, twenty repetitions, *i.e.* 200 outer folds). Within each fold, median imputation, FAMD and K-means ($K = 4$) were re-estimated on the training observations only; the test observations were then projected into the factor space learned on the training set and assigned to the nearest training centroid. Test assignments were compared with the full-cohort partition after optimal matching of the labels, and the out-of-sample agreement and the ARI were summarised by their mean and standard deviation across outer folds.

2.7. Low-Dimensional Representation

A UMAP projection (*Uniform Manifold Approximation and Projection*) [63] was

carried out for exclusively descriptive purposes, with 15 neighbours and a minimum distance of 0.1, from the FAMD factor coordinates. It is explicitly stressed that UMAP was not used to define the classes: the inter-cluster distances and the areas occupied in a UMAP plane are not quantitatively interpretable.

2.8. Sensitivity Analyses

Five sensitivity analyses were conducted, the first four of which were pre-specified. First, the classification was repeated on the complete observations for all quantitative variables only ($n = 144$), without any imputation, and the resulting partition compared with the main partition by the adjusted Rand index. Second, four different sets of classification variables were compared as to the quality and interpretability of the partitions produced. Third, the three-, five- and six-class solutions were examined. Fourth, the classification was repeated on 3 to 8 factor components. Fifth, the classification was repeated with the dual-coded specification, which includes both the origin of the suppuration and the grouped portal of entry.

The detailed clinical, therapeutic, microbiological and outcome characterisation (**Tables 2-6** and **Figures 1-5**), as well as the first four sensitivity analyses, concern the four phenotypes of the dual-coded specification. These phenotypes were retained for descriptive purposes because they carry the clinical labels used throughout the discussion; the primary non-redundant solution is reported through its validity indices, its reproducibility and its correspondence with these phenotypes.

2.9. Comparisons between Phenotypes

Quantitative variables were described by the median and the interquartile range, and compared by the Kruskal-Wallis test; qualitative variables by the count and percentage, compared by the Fisher-Freeman-Halton exact test (exact enumeration for two-category variables, Monte-Carlo simulation with 20,000 resamples for variables with more than two categories), several contingency tables containing sparse or zero cells. Given the exploratory nature of the analysis and the multiplicity of comparisons, the p-values are presented descriptively. The share of variance explained by the phenotype was estimated by the correlation ratio η^2 , derived from the one-way analysis-of-variance decomposition and reported as an effect size only; all p-values for quantitative variables are those of the Kruskal-Wallis test.

Two primary outcomes were analysed separately. The first was in-hospital mortality, available for all 150 observations. The second was functional status, defined as a GOS ≤ 3 at the last assessment among survivors with an available GOS (complete-case analysis). The eight missing GOS values were treated as missing, not as favourable, and two extreme scenarios—all missing values favourable, or all unfavourable—served as sensitivity analyses. The composite of death or GOS ≤ 3 was retained as an exploratory outcome only, since its two components refer to differ-

ent observation windows. Given the small number of deaths (16), the associations between phenotype and outcome were estimated by Firth-penalised logistic regression [64] [65], alone or adjusted for an admission GCS < 9 and for age. The confidence intervals of the proportions were computed by the Wilson method. The significance threshold was set at 0.05 two-sided, without correction for multiplicity given the exploratory nature of the study; the p-values must be read as descriptive.

The incremental prognostic value of the phenotype was then assessed formally, separately for each primary outcome, by comparing two nested logistic models: M0, including GCS and age, and M1, including GCS, age and phenotype. Discrimination (area under the ROC curve), overall accuracy (Brier score) and calibration (calibration intercept and slope) were corrected for optimism by bootstrap internal validation with 1000 replicates [66]: in each replicate, both models were refitted on the bootstrap sample and evaluated on the bootstrap sample and on the original sample, and the mean optimism was subtracted from the apparent performance. Given the small number of events, the models were kept deliberately parsimonious and the mortality estimates are regarded as imprecise.

2.10. Reproducibility

The analyses were carried out under Python 3.12 (Ubuntu 24.04), with *scikit-learn* 1.8, *prince*, *umap-learn*, *scipy* 1.17 and *statsmodels*. A single random seed (42) was fixed for the whole pipeline. The complete algorithm parameters are detailed in Supplementary **Table S1**: FAMD with 10 computed components and 15 iterations, of which the first 4 were retained; K-means with 500 initialisations (k-means++), default tolerance and maximum number of iterations, convergence achieved in all cases; Ward hierarchical classification on Euclidean distances; UMAP with 15 neighbours and a minimum distance of 0.1; random forest with 2000 trees without depth limitation. PAM was applied to Gower distances computed on the nineteen variables of the primary specification; the nested cross-validation used ten folds repeated twenty times; the optimism correction used 1000 bootstrap replicates. Binary variables were treated as two-category qualitative variables in the FAMD. Preprocessing comprised a check of duplicates, distributions and categories with a frequency below 3%, grouped as indicated above; no observation was excluded on the grounds of extreme values. The analysis code and the pseudonymised data are available from the corresponding author.

2.11. Ethical Considerations

The protocol was conducted in accordance with the Declaration of Helsinki. Data were extracted in a pseudonymised manner. The study was approved by the Institutional Ethics Committee of Bouaké University Hospital (reference CEI/CHU-BKE/119/2025, favourable opinion notified on 16 August 2026). Informed consent was obtained for the prospective inclusions (2025-2026); a waiver of consent was granted for the retrospective phase, the data having been pseudonymised.

3. Results

3.1. Population

One hundred and fifty observations corresponding to 148 patients were analysed. The median age was 26 years (IQR 14 - 50), with 40 observations (26.7%) younger than 15 years; the sex ratio was 2.3. The median delay between symptom onset and admission was 19 days (IQR 15 - 25). The median admission GCS was 13 (IQR 10 - 15), with 19 patients (12.7%) having a GCS < 9. Brain abscess was present in 93 cases (62%), subdural empyema in 27 (18%) and extradural empyema in 25 (17%)—these three lesion types being coded as non-exclusive presence indicators, an associated form being counted in each type concerned. In the epidemiological analysis of the same cohort, based on the dominant lesion type, brain abscess accounts for 73 observations (49%), extradural empyema for 20 (13%) and subdural empyema for 17 (11%), the multiple or associated forms there constituting a distinct category. In-hospital mortality was 10.7% (16/150; 95% CI 6.7 - 16.6) and unfavourable outcome 38.0% (57/150; 95% CI 30.6 - 46.0).

3.2. Primary Non-Redundant Classification

In the primary specification, from which the origin of the suppuration had been removed, a four-class solution was retained, with an overall silhouette coefficient of 0.296. Its concordance with the dual-coded partition was moderate (ARI 0.325). The frontal-interhemispheric sinogenic subdural empyema phenotype was the most robust, 28 of the 33 observations of phenotype 3 being recovered in a single group, whereas the post-traumatic and post-operative groupings of the dual-coded specification were redistributed according to lesion type and severity.

Partitioning around medoids on Gower distances, computed directly on the nineteen raw variables, identified four groups of 48, 36, 34 and 32 observations (Gower silhouette 0.312), in substantial concordance with the FAMD-K-means partition (ARI 0.671). In the repeated nested cross-validation of the complete pipeline, in which imputation, FAMD and K-means were re-estimated in each training set, the out-of-sample agreement after optimal label matching was 87.6% $\pm$ 4.8% and the mean ARI 0.79 $\pm$ 0.09 over 200 outer folds.

The detailed characterisation that follows (§ 3.3 to § 3.8) concerns the four phenotypes of the dual-coded specification, retained for descriptive purposes; their dependence on the weight given to the mechanism of infection is examined in § 3.9.

3.3. Factor Structure

In the dual-coded specification, the first four FAMD components concentrated respectively 11.9%, 10.7%, 8.1% and 7.2% of the inertia, *i.e.* 37.9% cumulatively (**Figure 1(A)**). The relatively diffuse inertia is expected with a large number of qualitative variables and does not indicate the absence of structure, as confirmed by the quality of the partitions obtained.

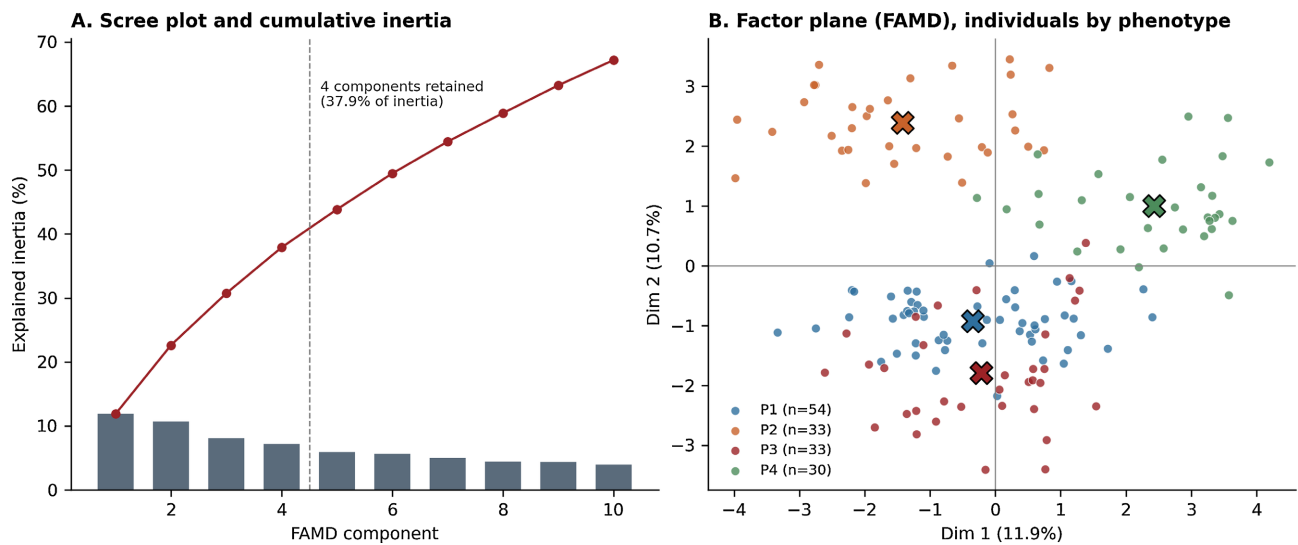


Figure 1. Factor analysis of mixed data. (A) Scree plot and cumulative inertia; four components were retained (37.9% of inertia). (B) Projection of the 150 observations onto the first factor plane, coloured by phenotype; the crosses represent the barycentres.

The first axis contrasted post-traumatic suppurations—with delayed presentation and voluminous collections—with early, pauci-lesional post-operative suppurations. The second axis separated ring-enhancing brain abscesses from frontal and interhemispheric subdural empyemas associated with hydrocephalus. The projection of the observations onto the first factor plane suggested an organisation compatible with four groupings, whose barycentres were separated, with residual overlap between some observations (**Figure 1(B)**).

3.4. Choice of the Number of Phenotypes

The internal validity indices of the dual-coded specification are detailed in **Table 2**. The silhouette coefficient had a local maximum at $k = 3$ (0.362) and $k = 4$ (0.358), clearly higher than at $k = 2$ (0.285), and rose again only beyond $k = 6$, to

Table 2. Internal validity indices by number of classes.

k	Silhouette (K-means)	Silhouette (Ward)	Calinski-Harabasz	Davies-Bouldin	ARI K-means/Ward	Mean Jaccard
2	0.285	0.285	46.3	1.289	1.000	0.69
3	0.362	0.361	64.2	1.110	0.958	0.94
4	0.358	0.356	71.5	1.056	0.958	0.91
5	0.345	0.324	73.2	0.907	0.779	0.88
6	0.349	0.335	71.4	0.976	0.659	0.80
7	0.359	0.344	72.1	0.965	—	—
8	0.370	0.354	75.8	0.979	—	—

ARI: adjusted Rand index. The retained solution is in bold. Stability assessed by the complete Hennig procedure, with 500 bootstrap resamples and re-assignment of all observations.

values ($k = 7$: 0.359; $k = 8$: 0.370) at which inter-algorithm concordance was no longer assessable. The Calinski-Harabasz index increased up to $k = 5$, and the Davies-Bouldin index was minimal at $k = 5$ (0.907) then rose. The gap statistic showed no local maximum before $k = 8$, a classic situation when the structure is continuous rather than strictly separated, and therefore did not discriminate between the solutions.

The per-class bootstrap stability, assessed by the complete Hennig procedure, was high for the four-class solution: mean Jaccard index of 0.91, with per-class values of 0.87 (phenotype 1), 1.00 (phenotype 2), 0.83 (phenotype 3) and 0.95 (phenotype 4). The four classes were thus above the 0.75 threshold defining a stable class, and three above the 0.85 threshold defining a highly stable class; phenotypes 2 and 4, the most mechanically homogeneous, reached near-perfect stability, whereas phenotype 3 was the most prone to marginal reclassification. Concordance between K-means and Ward reached 0.958 (ARI) for $k = 3$ as for $k = 4$, but fell to 0.779 as soon as $k = 5$, indicating that the structure becomes algorithm-dependent beyond four classes.

The four-phenotype solution was ultimately retained. At statistical quality equivalent to that of the three-class solution, it isolates post-operative suppuration—an entity of specific mechanism, microbial ecology and management—that the three-class solution merged with community-acquired abscess (ARI between the two solutions: 0.635). This choice explicitly favours clinical interpretability at equal statistical performance, in accordance with the recommendations on classification in medicine.

3.5. Description of the Four Phenotypes

The dendrogram of the Ward classification (**Figure 2**) and the UMAP projection (**Figure 3(A)**) illustrate the retained partition. The standardised profiles are shown in **Figure 4** and the detailed characteristics in **Table 3** and **Table 4**.

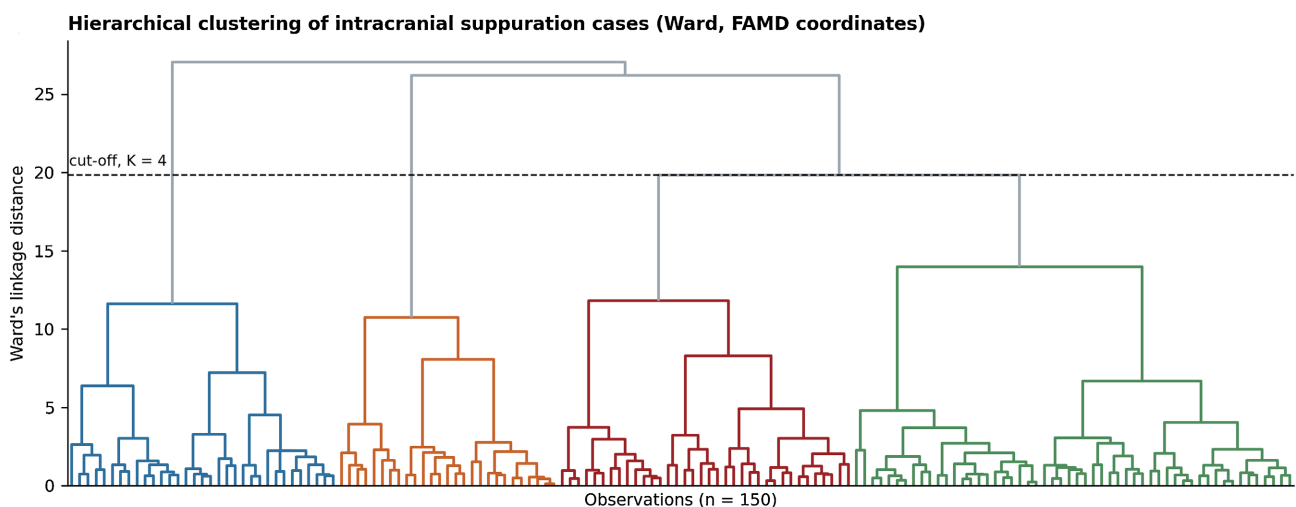


Figure 2. Dendrogram of the hierarchical agglomerative classification (Ward criterion, FAMD factor coordinates). The dashed line indicates the retained cut-off threshold ($K = 4$). Cophenetic correlation = 0.717.

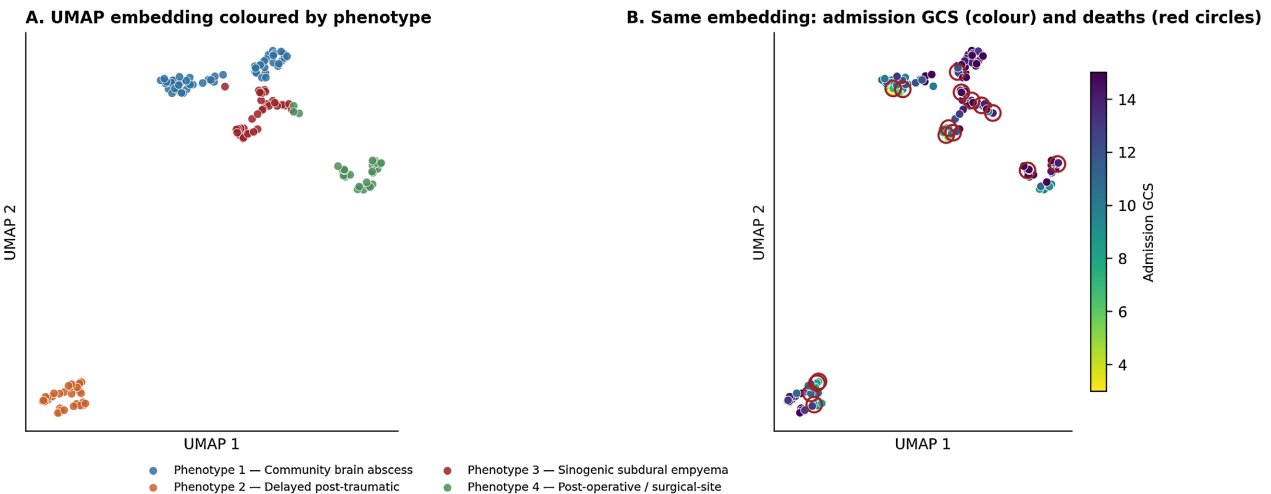


Figure 3. UMAP projection of the 150 observations. (A) Colouring by the phenotype identified by K-means. (B) Same projection coloured by the admission Glasgow score; the red circles identify deaths, whose distribution largely follows the consciousness gradient rather than phenotype membership. UMAP is used for descriptive purposes only; inter-cluster distances are not quantitatively interpretable.



Figure 4. Standardised phenotype profiles. Each cell represents the mean deviation of the phenotype from the cohort mean, expressed in standard deviations (z-score). The horizontal lines separate the quantitative, the clinical and lesional, and the therapeutic and outcome variables. The therapeutic, microbiological and outcome variables shown here did not participate in the construction of the phenotypes: they constitute a post-hoc characterisation.

Table 3. Baseline characteristics by phenotype (n = 150 observations).

Variable	P1 (n = 54)	P2 (n = 33)	P3 (n = 33)	P4 (n = 30)	Total	p
Epidemiology						
Age, years, median [IQR]	26 [15 - 53]	36 [18 - 51]	19 [9 - 51]	23 [10 - 46]	26 [14 - 50]	0.24
Age < 15 years, n (%)	12 (22)	6 (18)	11 (33)	11 (37)	40 (27)	—
Male sex, n (%)	30 (56)	26 (79)	23 (70)	25 (83)	104 (69)	0.033
Rural residence, n (%)	24 (44)	18 (55)	18 (55)	18 (60)	78 (52)	0.55
Referred patient, n (%)	28 (52)	16 (48)	20 (61)	13 (43)	77 (51)	0.57
Delay and mode of onset						
Symptom-to-admission delay, days, median [IQR]	19 [15 - 25]	29 [21 - 34]	17 [13 - 20]	15.5 [12 - 19]	19 [15 - 25]	<0.001
Progressive onset, n (%)	18 (33)	24 (73)	7 (21)	1 (3)	50 (33)	<0.001
Host factors						
Immunocompromise, n (%)	4 (7)	3 (9)	6 (18)	0 (0)	13 (9)	0.073
HIV infection, n (%)	0 (0)	0 (0)	3 (9)	0 (0)	3 (2)	0.027
Origin and portal of entry						
Community/spontaneous, n (%)	54 (100)	0 (0)	33 (100)	3 (10)	90 (60)	<0.001
Post-traumatic, n (%)	0 (0)	33 (100)	0 (0)	0 (0)	33 (22)	
Post-operative, n (%)	0 (0)	0 (0)	0 (0)	27 (90)	27 (18)	
ENT-dental portal of entry, n (%)	26 (48)	0 (0)	27 (82)	0 (0)	53 (35)	<0.001
Haematogenous/cryptogenic portal of entry, n (%)	26 (48)	0 (0)	2 (6)	0 (0)	28 (19)	
Clinical						
Admission GCS, median [IQR]	13 [10 - 15]	12 [10 - 14]	13 [10 - 14]	14 [12 - 15]	13 [10 - 15]	0.14
GCS < 9, n (%)	6 (11)	6 (18)	6 (18)	1 (3)	19 (13)	0.21
Fever, n (%)	33 (61)	20 (61)	23 (70)	18 (60)	94 (63)	0.83
Raised ICP syndrome, n (%)	39 (72)	24 (73)	28 (85)	16 (53)	107 (71)	0.057
Impaired consciousness, n (%)	30 (56)	19 (58)	16 (48)	8 (27)	73 (49)	0.048
Meningeal syndrome, n (%)	0 (0)	0 (0)	5 (15)	2 (7)	7 (5)	0.002
Focal motor deficit, n (%)	14 (26)	13 (39)	13 (39)	9 (30)	49 (33)	0.48
Seizures, n (%)	10 (19)	5 (15)	4 (12)	8 (27)	27 (18)	0.51
Scalp swelling/cellulitis, n (%)	3 (6)	0 (0)	9 (27)	0 (0)	12 (8)	<0.001
Imaging and lesional						
Brain abscess, n (%)	53 (98)	21 (64)	7 (21)	12 (40)	93 (62)	<0.001
Subdural empyema, n (%)	1 (2)	2 (6)	22 (67)	2 (7)	27 (18)	<0.001
Extradural empyema, n (%)	0 (0)	10 (30)	7 (21)	8 (27)	25 (17)	<0.001

Continued

Ventriculitis, n (%)	9 (17)	4 (12)	0 (0)	3 (10)	16 (11)	0.070
Scalp suppuration, n (%)	0 (0)	3 (9)	2 (6)	9 (30)	14 (9)	<0.001
Frontal topography, n (%)	18 (33)	9 (27)	20 (61)	11 (37)	58 (39)	<0.001
Interhemispheric topography, n (%)	0 (0)	0 (0)	9 (27)	0 (0)	9 (6)	
Bilateral/midline involvement, n (%)	0 (0)	0 (0)	9 (27)	0 (0)	9 (6)	<0.001
Multiple collections, n (%)	24 (44)	11 (33)	15 (45)	7 (23)	57 (38)	0.19
Maximum diameter, mm, median [IQR]	35.5 [30 - 42]	43 [30 - 49]	38 [34 - 43]	33 [27 - 40]	36 [30 - 44]	0.10
Mass effect, n (%)	40 (74)	26 (79)	30 (91)	14 (47)	110 (73)	0.001
Midline shift, mm, median [IQR]	1.7 [0 - 3.5]	2.7 [0 - 4.7]	3.3 [1.3 - 4.4]	0 [0 - 1.4]	1.8 [0 - 4.1]	0.012
Hydrocephalus, n (%)	5 (9)	0 (0)	10 (30)	1 (3)	16 (11)	<0.001
Ring enhancement, n (%)	45 (83)	18 (55)	6 (18)	12 (40)	81 (54)	<0.001
Laboratory						
Leucocytes, G/L, median [IQR]	15.1 [13.0 - 16.6]	14.6 [11.2 - 16.4]	15.0 [12.6 - 17.5]	14.1 [13.4 - 17.3]	14.7 [12.5 - 16.9]	0.56
CRP, mg/L, median [IQR]	95.8 [71 - 127]	91.2 [70 - 113]	90.1 [84 - 129]	94.3 [77 - 112]	92.1 [74 - 124]	0.80

p-values in bold in the p column are significant at the 0.05 threshold. The marked delay, host-factor, clinical and lesional variables served to build the classes; the others (sex, referral status, residence, raised-ICP syndrome, meningeal syndrome, HIV infection, ring enhancement, bilateral involvement, leucocytes, CRP) constitute a post-hoc characterisation. Topography was grouped into five categories for the purposes of the classification; the frontal frequency of this grouping (58; 39%) reduces to 51 (34%) in the reference topographic classification of the cohort (temporal 38; 25%; interhemispheric 9; 6%). Recoding the topography on this scheme and repeating the classification leave the partition unchanged (sensitivity analysis, § 3.9). The unit of analysis is the hospitalisation episode. Comparisons involving the variables used to build the partition (delay, GCS, origin, portal of entry, type of collection, mass effect, hydrocephalus, topography) are descriptive: their p-values do not constitute independent confirmatory tests, since the phenotypes were derived from these same variables. Only the variables not used for the classification (microbiology, management, outcome) permit an inferential reading. Given the exploratory nature of the analysis, interpretation favours effect sizes and confidence intervals over statistical significance alone.

Table 4. Management and microbiology by phenotype.

Variable	P1 (n = 54)	P2 (n = 33)	P3 (n = 33)	P4 (n = 30)	Total	p
Medical-surgical treatment, n (%)	46 (85)	27 (82)	33 (100)	19 (63)	125 (83)	<0.001
Craniotomy, n (%)	31 (57)	21 (64)	30 (91)	14 (47)	96 (64)	<0.001
External ventricular drainage, n (%)	3 (6)	0 (0)	8 (24)	0 (0)	11 (7)	<0.001
Admission-to-surgery delay, days, median [IQR]	2 [1 - 3]	2 [1 - 3]	2 [1 - 3]	2 [1 - 3]	2 [1 - 3]	0.60
Triple antibiotic therapy, n (%)	22 (41)	33 (100)	9 (27)	20 (67)	84 (56)	<0.001
Corticosteroid therapy, n (%)	12 (22)	11 (33)	11 (33)	2 (7)	36 (24)	0.030
Osmotherapy, n (%)	21 (39)	16 (48)	13 (39)	6 (20)	56 (37)	0.11
Duration of antibiotic therapy, weeks, median [IQR]	7 [6 - 8]	6 [6 - 7]	6 [6 - 7]	7 [6 - 8]	7 [6 - 7]	0.16
Pus sample obtained, n (%)	42 (78)	25 (76)	28 (85)	21 (70)	116 (77)	0.56

Continued

Positive culture (among sampled), n (%)	17/42 (40)	13/25 (52)	9/28 (32)	11/21 (52)	50/116 (43)	0.39
<i>Streptococcus</i> spp., n (%)	9 (53)	9 (69)	3 (33)	0 (0)	21 (42)	0.0003
<i>Staphylococcus aureus</i> , n (%)	6 (35)	3 (23)	4 (44)	4 (36)	17 (34)	
Gram-negative bacilli (incl. <i>P. aeruginosa</i>), n (%)	2 (12)	1 (8)	0 (0)	4 (36)	7 (14)	
Coagulase-negative staphylococcus, n (%)	0 (0)	0 (0)	0 (0)	3 (27)	3 (6)	
Polymicrobial flora, n (%)	0 (0)	0 (0)	2 (22)	0 (0)	2 (4)	
Sterile culture, n (%)	25 (60)	12 (48)	19 (68)	10 (48)	66 (57)	

For micro-organisms, percentages are reported relative to the positive cultures of each phenotype (P1: 17; P2: 13; P3: 9; P4: 11; total: 50), and not to the phenotype size; the organism categories are mutually exclusive, polymicrobial samples being classified as polymicrobial flora. The “Sterile culture” row is reported relative to the samples obtained (denominators P1: 42; P2: 25; P3: 28; P4: 21; total: 116), as is the “Positive culture” row.

Phenotype 1, community-acquired brain abscess of ENT-dental or cryptogenic origin (n = 54; 36.0%), was the most frequent phenotype. All suppurations were community-acquired or spontaneous, with an ENT-dental portal of entry in 48% of cases and haematogenous or cryptogenic in 48%. Brain abscess was present in 98% of observations and ring enhancement in 83%—the highest value of the four groups. Onset was subacute (65%), the median admission delay 19 days, the median age 26 years and the most balanced sex ratio of the cohort (1.2). Topography was diverse (frontal 33%, parietal 26%, temporal 20%, other 20%) and collections were multiple in 44% of cases. This phenotype had the lowest mortality (5.6%) and the smallest proportion of GCS < 9 (11%).

In phenotype 2, delayed post-traumatic suppuration (n = 33; 22.0%), all observations were of post-traumatic origin, secondary to an open head injury or a depressed fracture. This phenotype was distinguished by the longest symptom-to-admission delay of the cohort (median 29 days; IQR 21 - 34, versus 19 days overall), a progressive onset in 73% of cases and the largest lesions (median maximum diameter 43 mm; IQR 30 - 49). Male predominance was marked (sex ratio 3.7) and the median age the highest (36 years). Extradural empyema was present in 30% of cases and brain abscess in 64%, with a temporal predominance (48%). Triple antibiotic therapy was used in all patients of this group. Mortality was 12.1% and the follow-up duration the shortest (median 52 days); the respective share of a more frequent loss to follow-up, a more recent recruitment or different care pathways could not be determined.

Phenotype 3, frontal-interhemispheric sinogenic subdural empyema (n = 33; 22.0%), gathered the most severe forms in lesional terms. All suppurations were community-acquired, with an ENT portal of entry in 82% of cases. Subdural empyema was present in 67% of observations, frontal topography in 61% and interhemispheric in 27%; involvement was bilateral or midline in 27% of cases. Mass effect was present in 91% of observations, with the largest midline shift (median 3.3 mm), and hydrocephalus in 30%. Meningeal syndrome (15%) and scalp

swelling or cellulitis (27%) were significantly more frequent here. This phenotype had the highest proportion of immunocompromised patients (18%), including all the HIV-infected patients of the cohort (9%), and the lowest median age (19 years, with 33% of patients under 15 years). Management was systematically medical-surgical (100%), with craniotomy in 91% of cases and external ventricular drainage in 24%. It had the highest mortality (18.2%), the highest proportion of unfavourable outcome (48.5%) and the lowest median GOS (3).

Phenotype 4, post-operative and surgical-site suppuration ($n = 30$; 20.0%), grouped the infections occurring in the aftermath of a neurosurgical operation (post-operative origin in 90% of cases; post-operative portal of entry 87%, cutaneous/scalp 13%). It was characterised by the shortest admission delay (median 15.5 days), the mildest neurological presentation—impaired consciousness in only 27% of cases versus 49% overall, GCS < 9 in 3%—and the lowest lesion burden (mass effect 47%, midline shift 0 mm, multiple collections 23%). Scalp supuration was present in 30% of cases. It was the only phenotype for which surgical abstention was frequent (medical treatment alone 37%). Its microbiology was distinct, with the highest proportion of Gram-negative bacilli (36% of positive cultures, including the three *Pseudomonas aeruginosa* isolates of the cohort), the three coagulase-negative staphylococcus isolates and no streptococcus. It had the shortest hospital stay (median 9 days), the highest proportion of patients without sequelae (60%) and the highest median GOS (5).

3.6. Concordance between Algorithms and Stability

The cophenetic correlation of the Ward classification on factor coordinates was 0.717, indicating good fidelity of the dendrogram to the structure of the distances. Concordance between K-means and Ward reached an ARI of 0.958: only three observations were classified differently by the two algorithms.

The PAM classification on Gower distances, computed without dimensionality reduction, is reported in § 3.2: being based on the non-redundant variable set, it assesses the primary partition, with which it showed substantial concordance (ARI 0.671). Ward's criterion, which requires Euclidean distances, was applied only to the FAMD factor coordinates.

The overall silhouette coefficient of 0.358 lies within the range usually observed in clinical phenotyping on mixed data, where values of 0.30 to 0.40 are frequent and reflect partially permeable boundaries rather than an absence of structure [29]. The mean per-class silhouette coefficient was homogeneous across phenotypes: 0.378 (phenotype 1), 0.356 (phenotype 2), 0.332 (phenotype 3) and 0.354 (phenotype 4), for an overall value of 0.358. Only one observation out of 150 (0.7%), belonging to phenotype 4, had a negative silhouette, *i.e.* an ambiguous assignment with respect to the geometry of the classes.

Descriptively, a random forest re-assigned the observations to their phenotype with an out-of-bag error of 4.7% and a stratified cross-validation accuracy of 95.3% (balanced accuracy 94.5%). Internal separability was high for the four

classes (one-versus-all AUC from 0.994 to 1.000; macro AUC 0.998) and calibration was satisfactory (multiclass Brier score 0.087; per-class scores from 0.000 to 0.040). The confusion matrix (**Table 5**) showed that the seven misclassifications were distributed exclusively between phenotype 3, four observations of which were assigned to phenotype 1, and phenotype 4, three observations of which were assigned to phenotype 3; phenotypes 1 and 2 were recognised without error (sensitivity 100%). This configuration is consistent with the bootstrap stability values, the same classes proving the most permeable by both approaches. The most discriminating variables were the origin of the suppuration, the portal of entry, the presence of a brain abscess, the presence of a subdural empyema, then the symptom-to-admission delay (Supplementary **Figure S1**). No single variable alone explained the partition, which rules out the hypothesis of a trivial classification reproducing a single input variable. The symptom-to-admission delay was the only quantitative variable strongly structured by the phenotype ($\eta^2 = 0.32$; Kruskal-Wallis $p < 0.001$), far ahead of age ($\eta^2 = 0.03$; $p = 0.24$), GCS ($\eta^2 = 0.03$; $p = 0.14$) and lesion diameter ($\eta^2 = 0.04$; $p = 0.10$), the p -values being those of **Table 3**. The random-forest figures measure only the supervised restitution of labels fixed on the whole cohort and must not be read as a validation of the clustering; the reproducibility of the complete pipeline is given by the nested cross-validation (§ 3.2), whose out-of-sample agreement of 87.6% is the more realistic estimate.

Table 5. Per-class stability and supervised classifier performance for the dual-coded solution ($K = 4$).

Phenotype	n	Mean silhouette	Mean bootstrap Jaccard	Median Jaccard [IQR]	Sensitivity	Precision	AUC	Brier
P1, Community abscess	54	0.378	0.87	1.00 [0.93 - 1.00]	100%	93.1%	0.997	0.032
P2, Delayed post-traumatic	33	0.356	1.00	1.00 [1.00 - 1.00]	100%	100%	1.000	0.000
P3, Sinogenic empyema	33	0.332	0.83	0.94 [0.89 - 0.97]	87.9%	90.6%	0.994	0.040
P4, Post-operative	30	0.354	0.95	0.94 [0.93 - 0.97]	90.0%	100%	0.999	0.014
Overall	150	0.358	0.91	—	95.3%	95.5%	0.998	0.087

Confusion matrix (ten-fold stratified cross-validation; rows = observed phenotype, columns = predicted phenotype):

	P1	P2	P3	P4
P1	54	0	0	0
P2	0	33	0	0
P3	4	0	29	0
P4	0	0	3	27

The multiclass Brier score corresponds, for each observation, to the sum of the squared errors over the four class probabilities, averaged over all observations; the per-class scores are computed in a one-versus-all approach. Stability: 500 bootstrap resamples, Hennig procedure. Classifier: random forest, 2,000 trees, base variables only. AUC: area under the ROC curve in a one-versus-all approach. Brier: Brier score (the overall value is the multiclass score). Balanced accuracy: 94.5%.

3.7. Management and Microbiological Documentation

The therapeutic strategy differed significantly between phenotypes ($p < 0.001$): medical-surgical treatment concerned 100% of phenotype 3, 85% of phenotype 1, 82% of phenotype 2 and 63% of phenotype 4. Craniotomy ($p < 0.001$) and external ventricular drainage ($p < 0.001$) were concentrated in phenotype 3. The antibiotic regimen also differed ($p < 0.001$), triple therapy being systematic in phenotype 2 and in the minority in phenotype 3.

A pus sample was obtained in 77.3% of observations (116/150), with no difference between phenotypes ($p = 0.56$). Relative to the samples obtained, culture yield ranged from 32.1% (9/28, phenotype 3) to 52.4% (11/21, phenotype 4), without reaching the significance threshold ($p = 0.39$). Fifty cultures were positive in total.

Relative to positive cultures only, the nature of the organisms isolated—mutually exclusive categories, polymicrobial samples being classified as polymicrobial flora—differed markedly between phenotypes (Fisher-Freeman-Halton exact test by Monte-Carlo simulation, $p = 0.0003$). Streptococci accounted for 53% (9/17) of the positive cultures of phenotype 1 and 69% (9/13) of those of phenotype 2, but were entirely absent from phenotype 4 (0/11). Conversely, Gram-negative bacilli and coagulase-negative staphylococci were concentrated in phenotype 4. The only two polymicrobial floras belonged to phenotype 3. This distribution is consistent with the portals of entry defining each phenotype and provides an independent biological corroboration of the partition, the microbiological data not having been used to build the classes. It does not constitute an external validation in the strict sense, which would require an independent cohort.

3.8. Outcome

In-hospital mortality ranged from 5.6% (95% CI 1.9 - 15.1) in phenotype 1 to 18.2% (95% CI 8.6 - 34.4) in phenotype 3, passing through 10.0% in phenotype 4 and 12.1% in phenotype 2 (**Figure 5(A)**; exact test $p = 0.29$). The GOS score was available for 142 of the 150 observations; the eight missing values all concerned survivors and were distributed as 3, 2, 2 and 1 observations in phenotypes 1 to 4 ($p = 1.00$). Among the 126 survivors with an evaluable GOS, 41 (32.5%) had a $\text{GOS} \leq 3$: 16/48 (33.3%) in phenotype 1, 10/27 (37.0%) in phenotype 2, 10/25 (40.0%) in phenotype 3 and 5/26 (19.2%) in phenotype 4 (exact test $p = 0.38$). Under the extreme assumptions for the missing values, the proportion of unfavourable functional status among the 134 survivors ranged from 30.6% (all eight favourable) to 36.6% (all eight unfavourable), which did not modify the overall interpretation. The exploratory composite of death or $\text{GOS} \leq 3$ ranged from 26.7% (phenotype 4) to 48.5% (phenotype 3) (**Figure 5(B)**; $p = 0.30$).

On unadjusted Firth regression (**Table 6**), taking phenotype 1 as reference, the odds ratio of death was 3.48 (95% CI 0.86 - 14.04; $p = 0.080$) for phenotype 3, 2.24 (0.51 - 9.92) for phenotype 2 and 1.87 (0.39 - 9.02) for phenotype 4. After adjustment for a $\text{GCS} < 9$ and for age, these associations attenuated (phenotype 3: OR

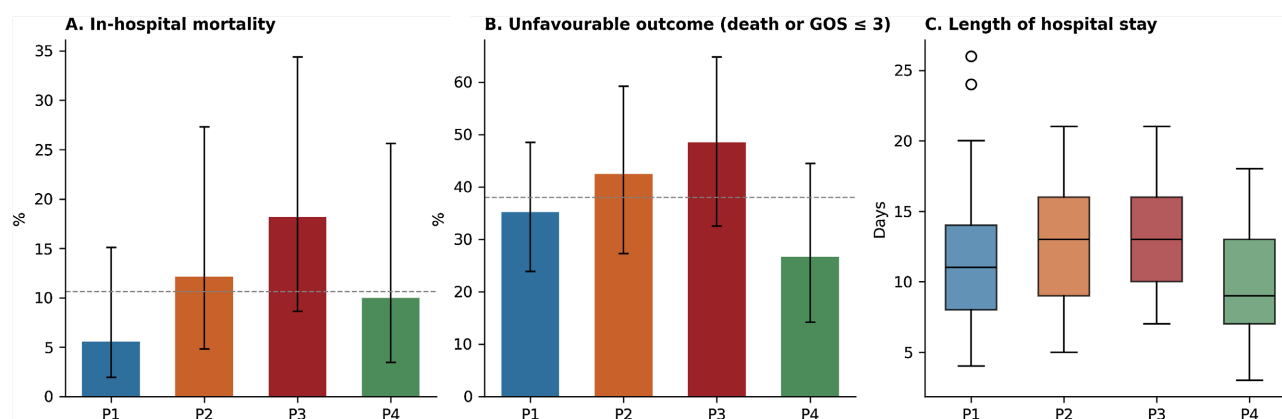


Figure 5. Outcome by phenotype. (A) In-hospital mortality (Wilson 95% CI). (B) Exploratory composite of death or GOS ≤ 3 (Wilson 95% CI). (C) Length of hospital stay. The dashed line (A), (B) represents the value for the whole cohort.

Table 6. Outcome by phenotype and Firth-penalised regression models.

Variable	P1 (n = 54)	P2 (n = 33)	P3 (n = 33)	P4 (n = 30)	Total	p
In-hospital death, n (%) [95% CI]	3 (5.6) [1.9 - 15.1]	4 (12.1) [4.8 - 27.3]	6 (18.2) [8.6 - 34.4]	3 (10.0) [3.5 - 25.6]	16 (10.7)	0.29
Unfavourable outcome, exploratory composite, n (%) [95% CI]	19 (35.2) [23.8 - 48.5]	14 (42.4) [27.2 - 59.2]	16 (48.5) [32.5 - 64.8]	8 (26.7) [14.2 - 44.4]	57 (38.0)	0.30
GOS missing, n (%)	3 (6)	2 (6)	2 (6)	1 (3)	8 (5)	1.00
GOS ≤ 3 among evaluable survivors, n/N (%)	16/48 (33.3)	10/27 (37.0)	10/25 (40.0)	5/26 (19.2)	41/126 (32.5)	0.38
GOS, median [IQR]	4 [3 - 5]	4 [3 - 5]	3 [2 - 4]	5 [3 - 5]	4 [3 - 5]	0.057
No sequelae, n (%)	17 (31)	10 (30)	8 (24)	18 (60)	53 (35)	0.018
Post-operative complication, n (%)	7 (13)	9 (27)	10 (30)	5 (17)	31 (21)	0.17
Surgical revision, n (%)	5 (9)	2 (6)	4 (12)	2 (7)	13 (9)	0.85
Length of hospital stay, days, median [IQR]	11 [8 - 14]	13 [9 - 16]	13 [10 - 16]	9 [7 - 13]	11 [8 - 15]	0.017

Exploratory composite: death or GOS ≤ 3 at the last assessment.

Firth models (reference: phenotype 1)

Model	Variable	OR	95% CI	p
Death, unadjusted	Phenotype 2	2.24	0.51 - 9.92	0.29
	Phenotype 3	3.48	0.86 - 14.04	0.080
	Phenotype 4	1.87	0.39 - 9.02	0.43
Death, adjusted	Phenotype 2	2.11	0.43 - 10.39	0.36
	Phenotype 3	3.18	0.77 - 13.04	0.108
	Phenotype 4	2.24	0.41 - 12.24	0.35
	GCS < 9	3.62	1.08 - 12.12	0.037
	Age (per year)	1.00	0.97 - 1.02	0.92

Continued

Unfavourable, unadjusted	Phenotype 2	1.36	0.56 - 3.30	0.50
	Phenotype 3	1.73	0.72 - 4.19	0.22
	Phenotype 4	0.67	0.25 - 1.79	0.43
Unfavourable, adjusted	Phenotype 2	1.15	0.43 - 3.10	0.78
	Phenotype 3	1.60	0.61 - 4.19	0.34
	Phenotype 4	0.82	0.29 - 2.33	0.72
	GCS < 9	13.71	3.39 - 55.50	<0.001
	Age (per year)	1.00	0.99 - 1.02	0.75

The death models (adjusted and unadjusted) were estimated by Firth-penalised logistic regression; the three unadjusted models of the composite criterion, in the absence of separation, by ordinary (unpenalised) logistic regression. The adjusted models include all phenotypes, GCS < 9 and age. The unit of analysis is the hospitalisation episode. GOS was available for 142/150 observations; the eight missing values, all in survivors (3, 2, 2 and 1 in phenotypes 1 to 4), were considered favourable in the exploratory composite only; the functional-status row is a complete-case analysis among evaluable survivors. Mortality is in-hospital whereas GOS is collected at the last available assessment: the composite criterion therefore combines two distinct time windows. The “No sequelae” row corresponds to the neurological-sequelae variable (53 observations, i.e. about 40% of the 134 survivors) and not to GOS level 5 (49 observations).

3.18; 0.77 - 13.04; $p = 0.108$) while GCS < 9 remained significantly associated with death (OR 3.62; 1.08 - 12.12; $p = 0.037$). For the composite criterion, after adjustment for GCS and age, no phenotype was significantly associated with unfavourable outcome (all $p > 0.34$), whereas GCS < 9 was strongly so (aOR 13.71; 95% CI 3.39 - 55.50; $p < 0.001$).

The formal assessment of incremental prognostic value showed a limited contribution of the phenotype (**Table 7**). For in-hospital mortality, the optimism-corrected AUC was 0.754 for model M0 (GCS and age) and 0.763 after the addition of phenotype (M1), with Brier scores of 0.091 and 0.090, calibration intercepts of -0.04 and -0.06 and calibration slopes of 0.91 and 0.84. For functional status among the 126 evaluable survivors, the corrected AUC was 0.794 for M0 and 0.803 for M1, with Brier scores of 0.181 and 0.179, intercepts of -0.02 and -0.03 and slopes of 0.94 and 0.89. The gain in discrimination was therefore 0.009 for each outcome, without meaningful improvement of overall accuracy or calibration; the lower calibration slope of M1 for mortality reflects the additional parameters estimated from only 16 events, and the mortality estimates remain imprecise.

By contrast, the length of hospital stay differed significantly between phenotypes ($p = 0.017$), from 9 days (IQR 7-13) in phenotype 4 to 13 days in phenotypes 2 and 3 (**Figure 5(C)**), as did the proportion of observations without sequelae, defined by the absence of neurological sequelae recorded at the last assessment (60% in phenotype 4 versus 24% in phenotype 3; exact test $p = 0.018$).

Because mortality was measured over the hospital period whereas GOS was collected at the last available assessment, whose timing varied between phenotypes (median follow-up 52 days in phenotype 2 versus 90 days in phenotype 1; $p = 0.055$), the two outcomes were analysed separately and the composite criterion is

Table 7. Incremental prognostic value of the phenotype beyond GCS and age (optimism-corrected by bootstrap, 1,000 replicates).

Outcome	Model	Corrected AUC	Corrected Brier score	Calibration intercept/slope
Mortality (n = 150; 16 deaths)	M0: GCS + age	0.754	0.091	−0.04/0.91
	M1: GCS + age + phenotype	0.763	0.090	−0.06/0.84
	Difference M1 - M0	+0.009	−0.001	—
Functional status (n = 126 evaluable survivors; 41 events)	M0: GCS + age	0.794	0.181	−0.02/0.94
	M1: GCS + age + phenotype	0.803	0.179	−0.03/0.89
	Difference M1 - M0	+0.009	−0.002	—

M0: logistic model including GCS and age; M1: the same model with phenotype added. Functional status: GOS ≤ 3 at the last assessment among the 126 survivors with an evaluable GOS (complete-case analysis). All values are corrected for optimism by bootstrap internal validation (1000 replicates); calibration intercept and slope are reported as intercept/slope.

reported as exploratory only.

The UMAP projection coloured by admission GCS and by the occurrence of death (**Figure 3(B)**) illustrates this result: deaths were not concentrated in a particular region of the phenotypic space, but followed the consciousness gradient, transverse to the four groups.

3.9. Sensitivity Analyses

With regard to imputation, the classification restricted to the 144 observations complete for all quantitative variables, performed without any imputation, restored a partition very close to the main partition (ARI 0.883): 139 of the 144 observations (96.5%) retained a homologous assignment.

As regards the number of classes, the five-class solution split phenotype 1 into two subgroups that no clinically relevant variable distinguished, with a drop in inter-algorithm concordance (ARI 0.779); the six-class solution further degraded stability; the three-class solution merged phenotypes 1 and 4 (ARI 0.635). Varying the number of retained factor components from 3 to 8 did not modify the four-class structure.

In the leave-one-variable-out analysis, the partition did not depend on any single quantitative variable: removing age, the symptom-to-admission delay or the Glasgow score left the partition strictly unchanged (ARI 1.000 in the three cases), and removing the maximum lesion diameter modified it marginally (ARI 0.972), as did removing impaired consciousness (ARI 0.972). Removing the structuring lesional variables had a more marked effect (without brain abscess: ARI 0.710; without subdural empyema: ARI 0.511; without topography: ARI 0.509).

The origin of the suppuration and the grouped portal of entry proved almost deterministically associated (Cramér's $V = 0.989$), the grouped portal of entry constituting a refinement of the origin. Their joint inclusion, which confers a double weight on the mechanism of onset, defines the dual-coded specification, now treated as a sensitivity analysis. Compared with it, the primary non-redundant solution showed a lower silhouette (0.296) and moderate concordance (ARI 0.325), and only the frontal-interhemispheric sinogenic phenotype was reconstituted almost completely (28 of the 33 observations of phenotype 3); the post-traumatic and post-operative phenotypes dissolved in favour of a grouping organised by lesion type and severity. Removing the portal of entry instead of the origin gave a similar picture (silhouette 0.294; ARI 0.291). This dependence is discussed in § 4.12.

In subpopulation analyses, the analysis restricted to adults restored the main partition ($n = 110$; ARI 0.923). The structure was, by contrast, less well restored after removal of the post-operative suppurations ($n = 123$; ARI 0.538) or in the subgroup of brain abscesses only ($n = 93$; ARI 0.460), which is expected since these restrictions remove by construction one of the axes of the partition. Finally, no period effect was demonstrated: the distribution of phenotypes did not differ between 2016-2020 and 2021-2026 ($p = 0.82$), the median year of admission was identical across the four groups ($p = 0.96$), and the classification conducted separately on each period restored the main partition (ARI 0.866 and 1.000). The post-operative phenotype is therefore not an artefact of the evolution of the centre's surgical activity.

In order to align the coding of topography with that of the reference epidemiological analysis of the same cohort (frontal 51; 34%; interhemispheric 9; 6%, the latter maintained as a distinct category), the topographic variable was recoded and the classification entirely repeated. The four-phenotype solution was preserved, with high concordance with the main partition (ARI ≈ 0.91) and the marginal redistribution of two observations (counts of 53, 34, 33 and 30 for phenotypes 1 to 4). Phenotype 3 remained frontal-interhemispheric and all conclusions remained unchanged: the partition therefore does not depend on the retained topographic coding scheme.

4. Discussion

4.1. Phenotypes and Anatomical Classification

To our knowledge, and following a structured literature search whose strategy is detailed in the appendix, this work constitutes the first application of an unsupervised classification to a cohort of intracranial suppurative infections, and the first study of automated phenotyping of this condition in an African series. Four phenotypes emerged spontaneously, without supervision and without the use of any therapeutic, microbiological or outcome variable. Their clinical coherence is high: each corresponds to an identifiable causal mechanism, a specific kinetics, a preferential lesion topography and a distinct microbial ecology. A contemporary work

(2026) described, for the *Streptococcus anginosus* group, compartmental phenotypes defined a priori on anatomical criteria and not by unsupervised learning, which leaves intact the originality of the present approach [67].

It is important to stress what this partition is not. It does not copy the anatomical classification: phenotype 2 mixes brain abscesses (64%) and extradural empyemas (30%) united by their post-traumatic mechanism, while brain abscesses are distributed across three of the four phenotypes. It is therefore the portals of entry and the delays, rather than the anatomical compartments, that principally structure the data, although this pre-eminence of the mechanism was partly constructed by its double weighting in the dual-coded specification (§ 4.12). Within this cohort, the dominant contribution of the symptom-to-admission delay ($\eta^2 = 0.32$) among the quantitative variables is in this respect one of the main results: in this care setting, the access-to-care trajectory is more discriminating than age, neurological severity or lesion size. This finding, difficult to formulate within a classical supervised analysis, illustrates the specific contribution of the unsupervised approach.

4.2. Robustness of the Classification

The classic criticism levelled at classification studies is that the identified classes could result from the investigators' methodological choices rather than from an intrinsic organisation of the data. Any partitioning algorithm produces classes, including on unstructured data [29]; the question is therefore never whether classes exist, but whether those produced survive reasonable perturbations of the method.

Several levels of validation were used. The first is inter-algorithm concordance: in the dual-coded specification, the K-means and Ward classifications on the same factor coordinates coincided almost exactly (ARI 0.958, three discordant observations), and in the primary specification, PAM on Gower distances—a geometry that borrows nothing from the FAMD—restored a substantially concordant partition (ARI 0.671). The second is stability under resampling: the complete Hennig procedure placed the four classes above the stability threshold, with a mean Jaccard index of 0.91. The third, and the most demanding, is the nested cross-validation of the complete pipeline, in which imputation, FAMD and K-means were re-estimated in each training set: the out-of-sample agreement reached 87.6% (mean ARI 0.79). This figure, lower than the 95.3% accuracy of the random forest, is the more realistic estimate of internal reproducibility, since the random forest only reproduces labels fixed on the whole cohort; the latter nonetheless indicates that the classes can be approximated by simple rules applied to the initial variables.

The fourth level is the most convincing because it is biological and not statistical. The microbiological data at no point participated in the construction of the classes; yet their distribution differed markedly between phenotypes ($p = 0.0003$), with a complete absence of streptococci and a concentration of Gram-negative bacilli and coagulase-negative staphylococci in the post-operative phenotype. A

purely artefactual partition would have no reason to reconstitute a bacterial ecology. This coherence constitutes an independent biological corroboration, in the sense that it rests on information that was not used for learning. It does not replace an external validation, which would require a distinct cohort.

4.3. Correspondence of the Phenotypes with the Literature

Each of the four phenotypes corresponds to entities described in isolation in the literature, which constitutes a form of content validation.

Phenotype 1 corresponds to the classic community-acquired brain abscess of Western and Asian series [1]-[3] [6] [8] [28] [68] [69], with its ENT-dental or cryptogenic portal of entry, its ring enhancement, its subacute onset and its relatively favourable prognosis. Contemporary series from high-resource referral centres find a comparable aetiological distribution [70]. The high proportion of cryptogenic forms (48%) is consistent with series of low microbiological yield and underlines the potential role of 16S rRNA sequencing [71]-[75].

Phenotype 2 covers post-traumatic suppurations, whose specificity—delayed presentation, voluminous collections, young-adult male predominance—is well documented in African and South-Asian series [15] [18]-[20]. The median delay of 29 days observed here far exceeds the values reported in high-resource series and reflects the conjunction of an initially neglected or incompletely debrided craniocerebral wound and a fragmented care pathway. This phenotype is, of all, the one whose very existence is a marker of the health system rather than of biology.

Phenotype 3 corresponds to sinogenic subdural empyema, an entity whose severity and frontal-interhemispheric topography have been known since the large South African series [15] [23] [24] and confirmed by Nigerian [25] [26] and East African works. The paediatric and adolescent predominance, the frequency of interhemispheric and bilateral involvement, the place of wide craniotomy—found here in 91% of patients, in accordance with Nathoo *et al.*'s demonstration on 699 cases [24]—and the highest mortality of the group are consistent with these data [23] [24] [76]. The higher frequency of immunocompromise and HIV infection in this phenotype deserves to be underlined in a high-prevalence context. The Tunisian series [19] finds a neighbouring configuration, with a smaller share of immunocompromise.

Phenotype 4 corresponds to surgical-site suppurations and post-operative empyemas [27], whose literature underlines the attenuated presentation, the frequency of the cutaneous portal of entry and the nosocomial ecology dominated by Gram-negative bacilli and coagulase-negative staphylococci. The recovery of this microbiological signature by a classification that used no microbiological variable constitutes the strongest independent corroboration argument of this work. This phenotype is also the one whose relative frequency should vary most from one centre to another, according to the volume of surgical activity.

The comparison with contemporary African series remains limited by their

methodological heterogeneity and by the scarcity of works published outside southern Africa and English-speaking West Africa [20]. The constitution of harmonised regional registries appears, in this respect, as a prerequisite for any rigorous comparison.

4.4. Justification of the Number of Phenotypes

The choice of the number of classes is the most debatable decision of any analysis of this type and deserves an explicit justification.

The two-class solution was clearly insufficient (silhouette 0.285): it simply opposed community forms to forms secondary to a breach, reducing the information to an already-known dichotomy. The three-class solution was statistically equivalent to the four-class one (silhouette 0.362 versus 0.358), but it merged post-operative suppuration with community-acquired abscess—two entities whose mechanism, microbial ecology, empirical antibiotic strategy and proportion of exclusively medical treatments differ profoundly. This merging would have masked the only phenotype for which anti-Gram-negative coverage is justified from the outset, *i.e.* precisely the result with the most direct therapeutic reach. The five-class solution, conversely, split phenotype 1 along an axis with no identifiable clinical correlate and saw inter-algorithm concordance fall from 0.958 to 0.779, a sign that the structure became dependent on the choice of algorithm rather than on the data.

The four-class solution is furthermore biologically plausible, in that it restores four pathophysiological routes of access of the organism to the intracranial contents: the ENT-dental or haematogenous route of contiguity, the traumatic breach, the direct sinogenic extension with subdural diffusion, and the operative inoculation. This correspondence must nonetheless be qualified: the double weighting of the mechanism in the dual-coded specification favoured it, and in the non-redundant specification only the sinogenic route remained clearly individualised. The convergence between the statistical and mechanistic criteria therefore supports $K = 4$ as a descriptive choice, not the existence of four natural entities.

4.5. Phenotypes and Prognosis

The mortality gradient observed between phenotypes—from 5.6% to 18.2%, *i.e.* a ratio of more than three—did not reach the significance threshold, and attenuated after adjustment for the initial level of consciousness. This observation calls for two, non-exclusive, readings.

The first is a power reading. With 16 deaths distributed across four groups, the study has no power to detect an odds ratio of the order of 3. For purely illustrative purposes, several hundred observations would probably have been necessary to detect, with satisfactory power, a mortality difference of this order; this post-hoc calculation does not constitute a formal estimate of the sample size required for a future validation. The wide confidence intervals (phenotype 3: 8.6% - 34.4%) are compatible with a clinically important effect as well as with the absence of an ef-

fect. The result must therefore be described as inconclusive rather than negative.

The second reading is complementary. The admission Glasgow score remained, after adjustment, significantly and strongly associated with unfavourable outcome (aOR 13.7; 95% CI 3.39 - 55.50; $p < 0.001$), and the UMAP projection shows that deaths were distributed according to a consciousness gradient transverse to the phenotypes rather than according to group membership. This configuration suggests that phenotype and severity constitute two largely orthogonal dimensions of ISIs: the phenotype describes *how* the disease arose and *how* it will be treated, whereas the GCS describes *where* the patient stands at the time of admission. A sinogenic empyema seen at a stage of preserved consciousness and a community-acquired abscess seen comatose do not have the same prognosis despite their respective phenotype. This dissociation, were it confirmed on larger numbers, would have a clear practical implication: the phenotypic classification should serve to organise care, not to predict the individual outcome, for which the GCS remains the reference instrument [6] [7] [28].

The formal assessment of incremental value supports this reading. Adding the phenotype to GCS and age raised the optimism-corrected AUC by only 0.009, for mortality as for functional status, without improvement of the Brier score or of calibration. Within the limits imposed by 16 deaths and 41 unfavourable functional outcomes, the phenotype therefore adds little individual prognostic information to GCS and age in this sample.

4.6. Implications for the Organisation of Care

The operational value of this classification lies in its early availability. The four phenotypes are identifiable from admission, from three elements accessible in any centre with a scanner: the mechanism (community, post-traumatic, post-operative), the presumed portal of entry and the type of collection on the first imaging study. The supervised restitution of the assignment by a random forest, together with the out-of-sample agreement of 87.6% of the complete pipeline, suggests that these phenotypes are describable by simple rules, which makes it conceivable to build a bedside orientation algorithm.

The practical consequences differ by phenotype. For the emergency physician or the first-line physician, recognising phenotype 3—child or adolescent, recent sinusitis, deficit and seizures, crescent-shaped frontal collection—mandates neurosurgical transfer without delay, these patients almost systematically requiring wide craniotomy and frequently a ventricular drain. For the neurosurgeon, the phenotype conditions the anticipation of the procedure: craniotomy and provision for a drain in phenotype 3, puncture-aspiration often sufficient in phenotype 1, revision of the surgical site with possible removal of hardware in phenotype 4. For the infectious-disease physician, the concentration of Gram-negative bacilli and the complete absence of streptococci among the eleven positive cultures of phenotype 4 suggest that the local nosocomial ecology should be explicitly taken into account in the choice of empirical antibiotic therapy for these patients. This

observation, based on small microbiological numbers, cannot alone modify a protocol and requires confirmation. Phenotype 2, finally, designates a population whose prognosis is decided upstream of the hospital, in the quality of the initial debridement of craniocerebral wounds and in the reduction of a delay of nearly a month: the levers there are largely pre-hospital and pertain to the organisation of emergency care rather than to neurosurgery.

In terms of triage, this classification could help distinguish two questions usually conflated: that of operative urgency, largely determined by the phenotype, and that of severity, determined by the GCS. A phenotype-3 patient with a preserved GCS remains a surgical emergency; a phenotype-4 patient with an altered GCS is a severe patient whose lesion does not necessarily require an immediate procedure. This dissociation has direct operational value in centres where the operating theatre is a constrained resource.

4.7. Phenotypes and Clinical Research

Therapeutic trials devoted to intracranial suppurative infections generally treat these infections as a homogeneous entity. Our results suggest that this assumption is simplistic: patients included in the same trial may belong to phenotypes whose mechanism and ecology differ to the point that the treatment response could diverge. An empirical antibiotic therapy optimal for a Gram-negative post-operative suppuration is not that of a community-acquired abscess of ENT origin; likewise, the optimal duration of antibiotic therapy or the threshold for surgical indication—a question on which the data comparing aspiration and excision remain, moreover, limited [77]—could depend on the underlying mechanism.

This heterogeneity could help explain the variability observed between published series. The precedent of acute respiratory distress syndrome is instructive: the reanalysis of randomised trials by latent subphenotypes revealed treatment effects of opposite direction, invisible in the overall analysis [32]. An analogous approach applied to ISIs, in the form of subgroup analyses pre-specified by phenotype in future trials and registries, would constitute a natural extension of this work.

4.8. Implications for African Health Systems

Sub-Saharan African health systems face specific constraints: scarcity of neurosurgeons, unequal access to imaging, prolonged transfer delays, restricted microbiological resources [10]–[14]. In such a context, the value of a classification tool depends as much on its frugality as on its performance.

From this point of view, the phenotypes described rest exclusively on information available from admission in any centre equipped with a scanner, without specialised laboratory work or molecular analysis. This frugality conditions their transferability. On a regional scale, such a classification could harmonise transfer criteria to referral centres, structure the planning of operative resources and provide a common language for the future African registries devoted to ISIs, today

non-existent. Phenotype 2, in particular, designates a quantifiable public-health objective: the reduction of a median delay of 29 days between the first symptoms and admission, which pertains to the organisation of primary care and the initial management of open head injuries far more than to neurosurgery.

4.9. Decision-Support Tools

The performance of the supervised classifier makes it technically conceivable to build an orientation tool integrated into the clinical record, proposing from the entry of the initial data the most probable phenotype and the associated implications: probability of craniotomy, possibility of a ventricular drain, empirical antibiotic spectrum, expected length of stay.

This prospect nonetheless calls for explicit caution. The model described here has undergone no external validation, which for the time being precludes any deployment [42] [43]. A classifier reproducing a partition obtained on the same cohort measures the internal separability of the classes, not their predictive validity in a new population. The contemporary reference frameworks—TRIPOD + AI for reporting [41], PROBAST then PROBAST + AI for risk of bias [42] [78], FUTURE-AI for the conditions of robustness, explainability and deployment [43], DECIDE-AI for early clinical evaluation [45]—define the necessary steps between an exploratory result such as this one and a usable tool. None of these steps has been taken at this stage. Such a tool would moreover be intended to assist clinical judgement and reduce the variability of practices, not to replace it, particularly in centres where neurosurgical expertise is limited—a context where the risk of over-reliance on an algorithm is highest.

4.10. Generalisation of the Results

The generalisability of this classification must be discussed with caution, distinguishing the existence of the phenotypes from their relative frequency.

The four phenotypes probably reflect universal pathophysiological routes and should, on that basis, be found in other geographical contexts. Their relative frequency, on the other hand, is liable to vary considerably according to local epidemiology, vaccination coverage, the incidence of open head injuries, the volume of neurosurgical activity and the organisation of the health system. A high-volume surgical centre would probably observe a higher proportion of post-operative phenotypes; a region with a high prevalence of untreated ENT infections would see community and sinogenic forms predominate; an efficient emergency-care system would reduce, or even eliminate, the delayed post-traumatic phenotype, whose median delay of 29 days is itself the product of a pathway failure.

Two characteristics observed here are, moreover, probably specific to the Ivorian context and must not be extrapolated: the higher frequency of HIV infection in phenotype 3 and the high proportion of cryptogenic forms, linked to the low microbiological yield. These reservations justify an external multicentre validation before any generalisation.

4.11. Conceptual Reach of the Results

Taken as a whole, these results fall within a general evolution of clinical research, which tends to move beyond inherited descriptive classifications in favour of classifications based on mechanisms and multidimensional profiles. This transition is established in sepsis [30], acute respiratory distress syndrome [31] [32], adult diabetes [33], asthma [34] and heart failure with preserved ejection fraction [35]. Our work suggests that neurosurgical infectious disease could undergo an analogous conceptual evolution.

The phenotypes described must be understood as latent clinical structures, descriptive and reproducible within this cohort, and not as causal pathological entities; other classification solutions could emerge in different populations. It would moreover be excessive to see them as a substitution for existing classifications. Anatomical classifications retain their own relevance: they determine the surgical procedure, and no phenotyping will ever dispense with knowing whether a collection is subdural or intraparenchymal. What these results suggest is a complementarity—the anatomical taxonomy answering the question of the procedure, the phenotypic classification those of the mechanism, the empirical antibiotic spectrum and the care trajectory. If these phenotypes were confirmed by international multicentre cohorts, they could usefully complement the way these infections are described, compared between centres and studied in trials.

4.12. Strengths and Limitations

The main strengths of this work are the exhaustiveness of the recruitment over a decade, the strict exclusion of therapeutic, microbiological and outcome variables from the classification phase, the assessment of robustness at several levels (inter-algorithm concordance in two geometries, per-class bootstrap stability, nested cross-validation of the complete pipeline, sensitivity analyses), the separate analysis of mortality and functional status, the formal assessment of incremental prognostic value and the independent biological corroboration by microbiology. It is furthermore, to our knowledge, the first African cohort of intracranial suppurative infections analysed by factor analysis of mixed data, unsupervised classification, UMAP projection, random forest and Firth-penalised regression.

Several limitations must nonetheless be underlined, the first of which is structural and conditions the interpretation of the whole.

The origin of the suppuration and the grouped portal of entry proved almost deterministically associated (Cramér's $V = 0.989$), the latter constituting a refinement of the former. Their joint inclusion, which was not an explicitly pre-specified weighting decision, conferred on the mechanism of onset a weight approximately double that of any other dimension. The primary analysis was therefore restricted to the non-redundant specification. Its concordance with the dual-coded partition was moderate (ARI 0.325), and only the frontal-interhemispheric sinogenic phenotype was then reconstituted almost entirely (28 of the 33 observations), the post-traumatic and post-operative phenotypes giving way to a group-

ing organised by lesion type and severity. The detailed characterisation of the four phenotypes presented here relies on the dual-coded partition, retained for descriptive purposes.

Two readings are possible and must be set out without premature arbitration. On the first, the pre-eminence of the mechanism in our results is in part constructed by the specification of the variables, and the assertion that pathophysiology structures the data better than anatomy would then be a consequence of the weight given to the mechanism rather than a property of the data. On the second, weighting the mechanism heavily is a legitimate clinical choice once it is declared, comparable to the domain weightings practised in other phenotyping works; origin and portal of entry are moreover not strictly synonymous, the latter distinguishing, within community forms, the ENT-dental, haematogenous/cryptogenic and cutaneous portals. In both readings, one fact remains: the frontal-interhemispheric sinogenic phenotype, the most severe and the one whose surgical reach is most direct, is the only one to resist all the specifications tested. The three other phenotypes must be considered conditional on the retained weighting, and this conditionality must be verified by external validation before any generalisation.

One hundred and fifty observations for nineteen classification variables (twenty in the dual-coded specification) remain below the ratios usually recommended [36] [37]. The prior dimensionality reduction attenuates this problem without cancelling it.

This work must be considered a derivation study. No independent cohort made it possible to verify that these phenotypes reconstitute elsewhere. The results must be read as hypothesis-generating.

The phenotypes reflect the population, the practices and the pathways of a single centre. Their relative frequency is not generalisable, and some observed characteristics are probably specific to the context.

The documentation of the retrospective observations is presumably less complete, which exposes to a non-quantifiable information bias.

With a positive culture in 43% of samples (33% of observations), the bacteriological data could not be integrated into the classification. No molecular sequencing was available, whereas 16S rRNA approaches reveal a substantial proportion of polymicrobial and anaerobic infections invisible on culture [71]-[75].

The variables describe the state on admission and the final outcome, without intermediate trajectory. A modelling of the outcome trajectories, which has made the richness of the work on sepsis, was not feasible.

Although the stability indices are high, the nature of the structure—discrete classes or poles of a continuum—is not settled by these data. The residual errors of the supervised classifier, concentrated at the boundaries of phenotypes 3 and 4, suggest transition zones.

Two patients each contribute two observations, introducing a minor, unmodelled dependence.

The imaging variables—mass effect, multiplicity, main topography, maximum

diameter, bilaterality—were extracted from routine radiology reports, without central re-reading or formal assessment of inter-observer reproducibility. No kappa coefficient or intraclass correlation is therefore available for these variables.

GOS was collected at the last available assessment and not at a uniform pre-specified time, with a follow-up duration differing between phenotypes; the composite criterion moreover combines in-hospital mortality and a functional status sometimes measured several months after discharge. A differential follow-up bias between phenotypes cannot be excluded; for this reason, mortality and functional status were analysed separately and the composite was retained as exploratory only.

The documented haematogenous and cryptogenic mechanisms were grouped into a single category owing to insufficient numbers, although they correspond to distinct situations; the distribution of the initial categories appears in the appendix.

No data from another health system were analysed; the reproducibility of the phenotypes outside the Ivorian context is unknown.

4.13. Perspectives

Four extensions appear as priorities. The first, indispensable, is external validation on multicentre African cohorts, the only way to reach the numbers necessary both to confirm the phenotypes and to test their own prognostic value. The second is the integration of microbiological variables obtained by 16S rRNA sequencing [71]–[75], which could make subphenotypes emerge that are currently masked by the high rate of sterile cultures. The third is the transition from discrete classes to continuous latent-variable models, better suited to a structure with transition zones. The fourth is the prospective testing of the central operational hypothesis of this work: does recognising the phenotype on admission effectively reduce the delays in operative decision-making and improve the adequacy of empirical antibiotic therapy?

5. Conclusion

The application of an unsupervised classification to an Ivorian cohort of 150 intracranial suppurative infections made four clinically coherent phenotypes emerge, stable under resampling and concordant between algorithms: community-acquired brain abscess of ENT-dental or cryptogenic origin, delayed post-traumatic suppuration, frontal-interhemispheric sinogenic subdural empyema and post-operative surgical-site suppuration. This classification, based exclusively on data available on admission, strongly structures the portal of entry, the lesion topography, the surgical requirement, the microbial ecology and the length of stay; the microbiological coherence of the classes, obtained without any bacteriological data participating in their construction, constitutes its most convincing corroboration. Once the dual coding of the mechanism was removed, however, only the frontal-interhemispheric sinogenic phenotype was consistently reproduced, the three others remaining conditional on the weight given to the mecha-

nism of onset. The classification adds, by contrast, little prognostic information to the initial Glasgow score and age (optimism-corrected gain in AUC of 0.009 for mortality and for functional status). In low-resource neurosurgery, the defensible contribution of this phenotyping is therefore the early orientation of care pathways, the anticipation of operative needs and the adjustment of the empirical antibiotic spectrum, while the prediction of the individual outcome must remain anchored on the assessment of consciousness on admission. External validation on multicentre African cohorts is the necessary prerequisite for any integration of this stratification into routine clinical decision-making. In its current state, this work must be regarded as hypothesis-generating and not as liable to modify practice.

Acknowledgements

The authors wish to express their deep gratitude to the entire staff of the Department of Neurosurgery of Bouaké University Hospital. We particularly thank the physicians, nurses, care assistants, administrative staff and all the collaborators who, through their daily commitment, their availability and their team spirit, contribute to patient care and to the scientific work of the department. This work is the fruit of a collective effort, and we wish, through these few lines, to pay tribute to the whole team for its valuable contribution.

Declarations

Ethical Approval

The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Ethics Committee of Bouaké University Hospital (reference CEI/CHU-BKE/119/2025, favourable opinion notified on 16 August 2026); informed consent for the prospective inclusions and waiver of consent for the retrospective phase (pseudonymised data).

Data and Code Availability

The pseudonymised data and the analysis code are available from the corresponding author on reasoned request, subject to the agreement of the ethics committee.

Related Publications

This work is a secondary analysis of a monocentric cohort of intracranial suppurative infections from Bouaké University Hospital (2016-2026) that has given rise to several studies sharing all or part of the same population: a main epidemiological article (clinical profile, management and prognostic factors of mortality) and other analyses addressing distinct questions (triage score, delays and care cascade, quality indicators, supervised machine learning, care trajectories). The present study is the only one to apply an unsupervised classification with a view to identifying clinical phenotypes and reproduces none of the results of these other

works. The population overlaps, as well as the differences in question, method and results, are declared to the editorial office.

Author Contributions

All authors contributed substantially to the design of the study, to the collection, analysis and interpretation of the data, as well as to the drafting or critical revision of the manuscript, and approved the final version.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Brouwer, M.C., Tunkel, A.R., McKhann, G.M. and van de Beek, D. (2014) Brain Abscess. *New England Journal of Medicine*, **371**, 447-456. <https://doi.org/10.1056/nejmra1301635>
- [2] Mathisen, G.E. and Johnson, J.P. (1997) Brain Abscess. *Clinical Infectious Diseases*, **25**, 763-779. <https://doi.org/10.1086/515541>
- [3] Brouwer, M.C. and van de Beek, D. (2017) Epidemiology, Diagnosis, and Treatment of Brain Abscesses. *Current Opinion in Infectious Diseases*, **30**, 129-134. <https://doi.org/10.1097/qco.0000000000000334>
- [4] Patel, K. and Clifford, D.B. (2014) Bacterial Brain Abscess. *The Neurohospitalist*, **4**, 196-204. <https://doi.org/10.1177/1941874414540684>
- [5] Bodilsen, J., Dalager-Pedersen, M., van de Beek, D., Brouwer, M.C. and Nielsen, H. (2020) Incidence and Mortality of Brain Abscess in Denmark: A Nationwide Population-Based Study. *Clinical Microbiology and Infection*, **26**, 95-100. <https://doi.org/10.1016/j.cmi.2019.05.016>
- [6] Brouwer, M.C., Coutinho, J.M. and van de Beek, D. (2014) Clinical Characteristics and Outcome of Brain Abscess: Systematic Review and Meta-Analysis. *Neurology*, **82**, 806-813. <https://doi.org/10.1212/wnl.0000000000000172>
- [7] Bodilsen, J., Duerlund, L.S., Mariager, T., Brandt, C.T., Petersen, P.T., Larsen, L., *et al.* (2022) Clinical Features and Prognostic Factors in Adults with Brain Abscess. *Brain*, **146**, 1637-1647. <https://doi.org/10.1093/brain/awac312>
- [8] Omland, L.H., Nielsen, H. and Bodilsen, J. (2024) Update and Approach to Patients with Brain Abscess. *Current Opinion in Infectious Diseases*, **37**, 211-219. <https://doi.org/10.1097/qco.0000000000001014>
- [9] Bodilsen, J., D'Alessandris, Q.G., Humphreys, H., Iro, M.A., Klein, M., Last, K., *et al.* (2024) European Society of Clinical Microbiology and Infectious Diseases Guidelines on Diagnosis and Treatment of Brain Abscess in Children and Adults. *Clinical Microbiology and Infection*, **30**, 66-89. <https://doi.org/10.1016/j.cmi.2023.08.016>
- [10] Meara, J.G., Leather, A.J.M., Hagander, L., Alkire, B.C., Alonso, N., Ameh, E.A., *et al.* (2015) Global Surgery 2030: Evidence and Solutions for Achieving Health, Welfare, and Economic Development. *The Lancet*, **386**, 569-624. [https://doi.org/10.1016/s0140-6736\(15\)60160-x](https://doi.org/10.1016/s0140-6736(15)60160-x)
- [11] Dewan, M.C., Rattani, A., Fieggen, G., Arraez, M.A., Servadei, F., Boop, F.A., *et al.* (2019) Global Neurosurgery: The Current Capacity and Deficit in the Provision of Essential Neurosurgical Care. Executive Summary of the Global Neurosurgery Ini-

- ative at the Program in Global Surgery and Social Change. *Journal of Neurosurgery*, **130**, 1055-1064. <https://doi.org/10.3171/2017.11.jns171500>
- [12] Corley, J., Lepard, J., Barthélemy, E., Ashby, J.L. and Park, K.B. (2019) Essential Neurosurgical Workforce Needed to Address Neurotrauma in Low- and Middle-Income Countries. *World Neurosurgery*, **123**, 295-299. <https://doi.org/10.1016/j.wneu.2018.12.042>
- [13] Park, K.B., Johnson, W.D. and Dempsey, R.J. (2016) Global Neurosurgery: The Unmet Need. *World Neurosurgery*, **88**, 32-35. <https://doi.org/10.1016/j.wneu.2015.12.048>
- [14] World Health Organization (2015) Strengthening Emergency and Essential Surgical Care and Anaesthesia as a Component of Universal Health Coverage (Resolution WHA68.15). World Health Organization.
- [15] Nathoo, N., Narotam, P.K., Nadvi, S. and van Dellen, J.R. (2012) Taming an Old Enemy: A Profile of Intracranial Suppuration. *World Neurosurgery*, **77**, 484-490. <https://doi.org/10.1016/j.wneu.2011.04.023>
- [16] Djientcheu, V.P., Mouafo, T.F., Esiene, A., Kamga, Y.N., Nguefack, S., Bello, F., *et al.* (2012) Intracranial Suppurations in the African Child: A Severe but Preventable Complication. *Child's Nervous System*, **29**, 119-123. <https://doi.org/10.1007/s00381-012-1930-6>
- [17] Miranda, H.A., Castellar Leones, S.M., Elzain, M.A. and Moscote-Salazar, L.R. (2013) Brain Abscess: Current Management. *Journal of Neurosciences in Rural Practice*, **04**, S67-S81. <https://doi.org/10.4103/0976-3147.116472>
- [18] Anwary, M.A. (2015) Intracranial Suppuration: Review of an 8-Year Experience at Umtata General Hospital and Nelson Mandela Academic Hospital, Eastern Cape, South Africa. *South African Medical Journal*, **105**, 584-588. <https://doi.org/10.7196/samjnew.7882>
- [19] Borni, M., Znazen, M., Kolsi, F. and Zaher Boudawara, M. (2022) Clinico-Radiological Assessment and Therapeutic Management of Intracranial Suppurations: Bicentric Series of 43 Cases with Literature Review. *Interdisciplinary Neurosurgery*, **27**, Article ID: 101393. <https://doi.org/10.1016/j.inat.2021.101393>
- [20] Nathoo, N., Nadvi, S.S., Narotam, P.K. and van Dellen, J.R. (2011) Brain Abscess: Management and Outcome Analysis of a Computed Tomography Era Experience with 973 Patients. *World Neurosurgery*, **75**, 716-726. <https://doi.org/10.1016/j.wneu.2010.11.043>
- [21] Sonnevile, R., Ruimy, R., Benzonana, N., Riffaud, L., Carsin, A., Tadié, J.M., *et al.* (2017) An Update on Bacterial Brain Abscess in Immunocompetent Patients. *Clinical Microbiology and Infection*, **23**, 614-620. <https://doi.org/10.1016/j.cmi.2017.05.004>
- [22] Menon, S., Bharadwaj, R., Chowdhary, A., Kaundinya, D.V. and Palande, D.A. (2008) Current Epidemiology of Intracranial Abscesses: A Prospective 5 Year Study. *Journal of Medical Microbiology*, **57**, 1259-1268. <https://doi.org/10.1099/jmm.0.47814-0>
- [23] Nathoo, N., Nadvi, S.S., van Dellen, J.R. and Gouws, E. (1999) Intracranial Subdural Empyemas in the Era of Computed Tomography: A Review of 699 Cases. *Neurosurgery*, **44**, 529-535. <https://doi.org/10.1097/00006123-199903000-00055>
- [24] Nathoo, N., Nadvi, S.S., Gouws, E. and van Dellen, J.R. (2001) Craniotomy Improves Outcomes for Cranial Subdural Empyemas: Computed Tomography-Era Experience with 699 Patients. *Neurosurgery*, **49**, 872-878. <https://doi.org/10.1227/00006123-200110000-00017>
- [25] Chikani, M., Mezue, W., Okorie, E., Mbachu, C., Ndubisi, C. and Chikani, U. (2017)

- Subdural Empyema: Clinical Presentations and Management Options for an Uncommon Neurosurgical Emergency in a Developing Country. *Nigerian Journal of Clinical Practice*, **20**, 1221-1225. https://doi.org/10.4103/njcp.njcp_340_16
- [26] Jolayemi, E.O., Bankole, O.B., Ojo, O.A., Bamigboye, B., Adebayo, B.O., Arekhandia, B.J., et al. (2022) Contemporary Management of Intracranial Subdural Empyema. *Journal of West African College of Surgeons*, **12**, 56-63. https://doi.org/10.4103/jwas.jwas_127_22
- [27] Hlavin, M.L., Kaminski, H.J., Fenstermaker, R.A. and White, R.J. (1994) Intracranial Suppuration: A Modern Decade of Postoperative Subdural Empyema and Epidural Abscess. *Neurosurgery*, **34**, 974-981. <https://doi.org/10.1227/00006123-199406000-00004>
- [28] Tseng, J.H. and Tseng, M.Y. (2006) Brain Abscess in 142 Patients: Factors Influencing Outcome and Mortality. *Surgical Neurology*, **65**, 557-562. <https://doi.org/10.1016/j.surneu.2005.09.029>
- [29] Loftus, T.J., Shickel, B., Balch, J.A., Tighe, P.J., Abbott, K.L., Fazzone, B., et al. (2022) Phenotype Clustering in Health Care: A Narrative Review for Clinicians. *Frontiers in Artificial Intelligence*, **5**, Article 842306. <https://doi.org/10.3389/frai.2022.842306>
- [30] Seymour, C.W., Kennedy, J.N., Wang, S., Chang, C.H., Elliott, C.F., Xu, Z., et al. (2019) Derivation, Validation, and Potential Treatment Implications of Novel Clinical Phenotypes for Sepsis. *JAMA*, **321**, 2003-2017. <https://doi.org/10.1001/jama.2019.5791>
- [31] Calfee, C.S., Delucchi, K., Parsons, P.E., Thompson, B.T., Ware, L.B. and Matthay, M.A. (2014) Subphenotypes in Acute Respiratory Distress Syndrome: Latent Class Analysis of Data from Two Randomised Controlled Trials. *The Lancet Respiratory Medicine*, **2**, 611-620. [https://doi.org/10.1016/s2213-2600\(14\)70097-9](https://doi.org/10.1016/s2213-2600(14)70097-9)
- [32] Famous, K.R., Delucchi, K., Ware, L.B., Kangelaris, K.N., Liu, K.D., Thompson, B.T., et al. (2017) Acute Respiratory Distress Syndrome Subphenotypes Respond Differently to Randomized Fluid Management Strategy. *American Journal of Respiratory and Critical Care Medicine*, **195**, 331-338. <https://doi.org/10.1164/rccm.201603-0645oc>
- [33] Ahlqvist, E., Storm, P., Käräjämäki, A., Martinell, M., Dorkhan, M., Carlsson, A., et al. (2018) Novel Subgroups of Adult-Onset Diabetes and Their Association with Outcomes: A Data-Driven Cluster Analysis of Six Variables. *The Lancet Diabetes & Endocrinology*, **6**, 361-369. [https://doi.org/10.1016/s2213-8587\(18\)30051-2](https://doi.org/10.1016/s2213-8587(18)30051-2)
- [34] Lötval, J., Akdis, C.A., Bacharier, L.B., Bjermer, L., Casale, T.B., Custovic, A., et al. (2011) Asthma Endotypes: A New Approach to Classification of Disease Entities within the Asthma Syndrome. *Journal of Allergy and Clinical Immunology*, **127**, 355-360. <https://doi.org/10.1016/j.jaci.2010.11.037>
- [35] Shah, S.J., Katz, D.H., Selvaraj, S., Burke, M.A., Yancy, C.W., Gheorghiad, M., et al. (2015) Phenomapping for Novel Classification of Heart Failure with Preserved Ejection Fraction. *Circulation*, **131**, 269-279. <https://doi.org/10.1161/circulationaha.114.010637>
- [36] Kaufman, L. and Rousseeuw, P.J. (1990) Finding Groups in Data: An Introduction to Cluster Analysis. Wiley. <https://doi.org/10.1002/9780470316801>
- [37] Hastie, T., Tibshirani, R. and Friedman, J. (2009) The Elements of Statistical Learning. Springer. <https://doi.org/10.1007/978-0-387-84858-7>
- [38] Riley, R.D., Ensor, J., Snell, K.I.E., Harrell, F.E., Martin, G.P., Reitsma, J.B., et al. (2020) Calculating the Sample Size Required for Developing a Clinical Prediction Model. *BMJ*, **368**, m441. <https://doi.org/10.1136/bmj.m441>

- [39] von Elm, E., Altman, D.G., Egger, M., Pocock, S.J., Gøtzsche, P.C. and Vandenbroucke, J.P. (2007) The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies. *The Lancet*, **370**, 1453-1457. [https://doi.org/10.1016/s0140-6736\(07\)61602-x](https://doi.org/10.1016/s0140-6736(07)61602-x)
- [40] Vandenbroucke, J.P., von Elm, E., Altman, D.G., Gøtzsche, P.C., Mulrow, C.D., Pocock, S.J., *et al.* (2007) Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and Elaboration. *PLOS Medicine*, **4**, e297. <https://doi.org/10.1371/journal.pmed.0040297>
- [41] Collins, G.S., Moons, K.G.M., Dhiman, P., Riley, R.D., Beam, A.L., Van Calster, B., *et al.* (2024) TRIPOD + AI Statement: Updated Guidance for Reporting Clinical Prediction Models That Use Regression or Machine Learning Methods. *BMJ*, **385**, e078378. <https://doi.org/10.1136/bmj-2023-078378>
- [42] Moons, K.G.M., Damen, J.A.A., Kaul, T., Hooft, L., Andaur Navarro, C., Dhiman, P., *et al.* (2025) PROBAST + AI: An Updated Quality, Risk of Bias, and Applicability Assessment Tool for Prediction Models Using Regression or Artificial Intelligence Methods. *BMJ*, **388**, e082505. <https://doi.org/10.1136/bmj-2024-082505>
- [43] Lekadir, K., Frangi, A.F., Porras, A.R., FUTURE-AI Consortium, *et al.* (2025) FUTURE-AI: International Consensus Guideline for Trustworthy and Deployable Artificial Intelligence in Healthcare. *BMJ*, **388**, e081554.
- [44] Norgeot, B., Quer, G., Beaulieu-Jones, B.K., Torkamani, A., Dias, R., Gianfrancesco, M., *et al.* (2020) Minimum Information about Clinical Artificial Intelligence Modeling: The MI-CLAIM Checklist. *Nature Medicine*, **26**, 1320-1324. <https://doi.org/10.1038/s41591-020-1041-y>
- [45] Vasey, B., Nagendran, M., Campbell, B., Clifton, D.A., Collins, G.S., Denaxas, S., *et al.* (2022) Reporting Guideline for the Early-Stage Clinical Evaluation of Decision Support Systems Driven by Artificial Intelligence: Decide-AI. *Nature Medicine*, **28**, 924-933. <https://doi.org/10.1038/s41591-022-01772-9>
- [46] Liu, X., Cruz Rivera, S., Moher, D., Calvert, M.J., Denniston, A.K., Chan, A., *et al.* (2020) Reporting Guidelines for Clinical Trial Reports for Interventions Involving Artificial Intelligence: The CONSORT-AI Extension. *Nature Medicine*, **26**, 1364-1374. <https://doi.org/10.1038/s41591-020-1034-x>
- [47] Cruz Rivera, S., Liu, X., Chan, A., Denniston, A.K., Calvert, M.J., Darzi, A., *et al.* (2020) Guidelines for Clinical Trial Protocols for Interventions Involving Artificial Intelligence: The SPIRIT-AI Extension. *Nature Medicine*, **26**, 1351-1363. <https://doi.org/10.1038/s41591-020-1037-7>
- [48] Buuren, S.V. and Groothuis-Oudshoorn, K. (2011) Mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, **45**, 1-67. <https://doi.org/10.18637/jss.v045.i03>
- [49] Stekhoven, D.J. and Bühlmann, P. (2011) MissForest—Non-Parametric Missing Value Imputation for Mixed-Type Data. *Bioinformatics*, **28**, 112-118. <https://doi.org/10.1093/bioinformatics/btr597>
- [50] Sterne, J.A.C., White, I.R., Carlin, J.B., Spratt, M., Royston, P., Kenward, M.G., *et al.* (2009) Multiple Imputation for Missing Data in Epidemiological and Clinical Research: Potential and Pitfalls. *BMJ*, **338**, b2393. <https://doi.org/10.1136/bmj.b2393>
- [51] Pagès, J. (2004) Analyse Factorielle de données Mixtes. *Revue de Statistique Appliquée*, **52**, 93-111.
- [52] Ward, J.H. (1963) Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, **58**, 236-244. <https://doi.org/10.1080/01621459.1963.10500845>

- [53] Kaufman, L. and Rousseeuw, P.J. (1987) Statistical Data Analysis Based on the L1-Norm and Related Methods. *Proceedings of the First International Conference on Statistical Data Analysis Based on the L1-Norm and Related Methods*, Neuchâtel, 31 August-4 September 1987, 405-416. <https://ci.nii.ac.jp/ncid/BA01083164>
- [54] Gower, J.C. (1971) A General Coefficient of Similarity and Some of Its Properties. *Biometrics*, **27**, 857-871. <https://doi.org/10.2307/2528823>
- [55] Hubert, L. and Arabie, P. (1985) Comparing Partitions. *Journal of Classification*, **2**, 193-218. <https://doi.org/10.1007/bf01908075>
- [56] Rousseeuw, P.J. (1987) Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *Journal of Computational and Applied Mathematics*, **20**, 53-65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- [57] Caliński, T. and Harabasz, J. (1974) A Dendrite Method for Cluster Analysis. *Communications in Statistics*, **3**, 1-27. <https://doi.org/10.1080/03610927408827101>
- [58] Davies, D.L. and Bouldin, D.W. (1979) A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **1**, 224-227. <https://doi.org/10.1109/tpami.1979.4766909>
- [59] Tibshirani, R., Walther, G. and Hastie, T. (2001) Estimating the Number of Clusters in a Data Set via the Gap Statistic. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **63**, 411-423. <https://doi.org/10.1111/1467-9868.00293>
- [60] Hennig, C. (2007) Cluster-Wise Assessment of Cluster Stability. *Computational Statistics & Data Analysis*, **52**, 258-271. <https://doi.org/10.1016/j.csda.2006.11.025>
- [61] Breiman, L. (2001) Random Forests. *Machine Learning*, **45**, 5-32. <https://doi.org/10.1023/a:1010933404324>
- [62] Brier, G.W. (1950) Verification of Forecasts Expressed in Terms of Probability. *Monthly Weather Review*, **78**, 1-3. [https://doi.org/10.1175/1520-0493\(1950\)078<0001:vofeit>2.0.co;2](https://doi.org/10.1175/1520-0493(1950)078<0001:vofeit>2.0.co;2)
- [63] McInnes, L., Healy, J. and Melville, J. (2018) UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv: 1802.03426.
- [64] Firth, D. (1993) Bias Reduction of Maximum Likelihood Estimates. *Biometrika*, **80**, 27-38. <https://doi.org/10.1093/biomet/80.1.27>
- [65] Heinze, G. and Schemper, M. (2002) A Solution to the Problem of Separation in Logistic Regression. *Statistics in Medicine*, **21**, 2409-2419. <https://doi.org/10.1002/sim.1047>
- [66] Harrell, F.E., Lee, K.L. and Mark, D.B. (1996) Multivariable Prognostic Models: Issues in Developing Models, Evaluating Assumptions and Adequacy, and Measuring and Reducing Errors. *Statistics in Medicine*, **15**, 361-387. [https://doi.org/10.1002/\(sici\)1097-0258\(19960229\)15:4<361::aid-sim168>3.0.co;2-4](https://doi.org/10.1002/(sici)1097-0258(19960229)15:4<361::aid-sim168>3.0.co;2-4)
- [67] Feng, J., Zhang, B., Lu, W., Liu, H., Liu, Y., Zou, Y., et al. (2026) Streptococcus Anginosus Group Brain Abscesses and Subdural Empyemas: Exploratory Compartment-Specific Phenotypes and Discharge Outcomes in a Single-Center Retrospective Cohort Study. *BMC Infectious Diseases*, **26**, Article No. 1560. <https://doi.org/10.1186/s12879-026-13998-w>
- [68] Helweg-Larsen, J., Astradsson, A., Richhall, H., Erdal, J., Laursen, A. and Brennum, J. (2012) Pyogenic Brain Abscess, a 15 Year Survey. *BMC Infectious Diseases*, **12**, Article No. 332. <https://doi.org/10.1186/1471-2334-12-332>
- [69] Xiao, F., Tseng, M., Teng, L., Tseng, H. and Tsai, J. (2005) Brain Abscess: Clinical Experience and Analysis of Prognostic Factors. *Surgical Neurology*, **63**, 442-449. <https://doi.org/10.1016/j.surneu.2004.08.093>

- [70] Evans, T.J., Jawad, S., Kalyal, N., Nadarajah, A., Amarouche, M., Stapleton, S., *et al.* (2022) Retrospective Review of the Epidemiology, Microbiology, Management and Outcomes of Intra-Cranial Abscesses at a Neurosurgical Tertiary Referral Centre, 2018-2020. *Annals of Clinical Microbiology and Antimicrobials*, **21**, Article No. 58. <https://doi.org/10.1186/s12941-022-00550-2>
- [71] Al Masalma, M., Lonjon, M., Richet, H., Dufour, H., Roche, P.H., Drancourt, M., *et al.* (2011) Metagenomic Analysis of Brain Abscesses Identifies Specific Bacterial Associations. *Clinical Infectious Diseases*, **54**, 202-210. <https://doi.org/10.1093/cid/cir797>
- [72] Kommedal, Ø., Wilhelmsen, M.T., Skrede, S., Meisal, R., Jakovljevic, A., Gaustad, P., *et al.* (2014) Massive Parallel Sequencing Provides New Perspectives on Bacterial Brain Abscesses. *Journal of Clinical Microbiology*, **52**, 1990-1997. <https://doi.org/10.1128/jcm.00346-14>
- [73] Stebner, A., Ensser, A., Geißdörfer, W., Bozhkov, Y. and Lang, R. (2021) Molecular Diagnosis of Polymicrobial Brain Abscesses with 16S-rDNA-Based Next-Generation Sequencing. *Clinical Microbiology and Infection*, **27**, 76-82. <https://doi.org/10.1016/j.cmi.2020.03.028>
- [74] John, D.V., Aryalakshmi, B., Deora, H., Purushottam, M., Raju, R., Mahadevan, A., *et al.* (2022) Identification of Microbial Agents in Culture-Negative Brain Abscess Samples by 16S/18S rRNA Gene PCR and Sequencing. *Tropical Biomedicine*, **39**, 489-498. <https://doi.org/10.47665/tb.39.4.002>
- [75] Son, H., Moon, J., Ha, E.J., Kim, N., Kim, E., Lee, H.S., *et al.* (2023) Identification of Bacterial Pathogens in Brain Abscesses by Metagenomic Approach Using Nanopore 16S Amplicon Sequencing. *Diagnostic Microbiology and Infectious Disease*, **107**, Article ID: 116041. <https://doi.org/10.1016/j.diagmicrobio.2023.116041>
- [76] Banerjee, A.D., Pandey, P., Devi, B.I., Sampath, S. and Chandramouli, B.A. (2009) Pediatric Supratentorial Subdural Empyemas: A Retrospective Analysis of 65 Cases. *Pediatric Neurosurgery*, **45**, 11-18. <https://doi.org/10.1159/000202619>
- [77] Lannon, M., Trivedi, A., Martyniuk, A., Farrokhyar, F. and Sharma, S. (2022) Surgical Aspiration versus Excision for Intraparenchymal Abscess: A Systematic Review and Meta-Analysis. *British Journal of Neurosurgery*, **36**, 743-749. <https://doi.org/10.1080/02688697.2022.2118231>
- [78] Wolff, R.F., Moons, K.G.M., Riley, R.D., Whiting, P.F., Westwood, M., Collins, G.S., *et al.* (2019) PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Annals of Internal Medicine*, **170**, 51-58. <https://doi.org/10.7326/m18-1376>

Supplementary Material

Table S1. Complete algorithm parameters.

Step	Library	Parameters
Imputation	pandas	Cohort median (maximum lesion diameter, 6 values)
FAMD	prince (FAMD)	n_components = 10 (4 retained), n_iter = 15, random_state = 42; binaries treated as 2-category qualitative
K-means	scikit-learn (KMeans)	n_clusters = 4, n_init = 500, init = k-means++, default tol and max_iter, random_state = 42
Ward	scipy (linkage/fcluster)	method = 'ward', metric = Euclidean, criterion = 'maxclust'
Gower	in-house implementation	19 raw variables of the primary specification, without dimensionality reduction; PAM, K = 4
Stability	scikit-learn + numpy	500 bootstrap resamples, Hennig procedure with complete re-assignment
UMAP	umap-learn	n_neighbors = 15, min_dist = 0.1, metric = Euclidean, random_state = 42
Random forest	scikit-learn	n_estimators = 2000, oob_score = True, unlimited depth, random_state = 42; 10-fold stratified cross-validation
Firth	statsmodels + in-house implementation	Jeffreys penalised likelihood; Wald 95% CI
Exact tests	scipy	Fisher-Freeman-Halton: exact enumeration for two-category variables; Monte-Carlo simulation (20,000 resamples) for variables with more than two categories
Nested CV	Same libraries as above	10 folds × 20 repetitions; imputation, FAMD and K-means re-estimated in each training set; optimal label matching
Incremental value	Same libraries as above	M0 (GCS, age) vs M1 (GCS, age, phenotype); bootstrap optimism correction, 1000 replicates

Single random seed (42) for the whole pipeline. Python 3.12, Ubuntu 24.04.

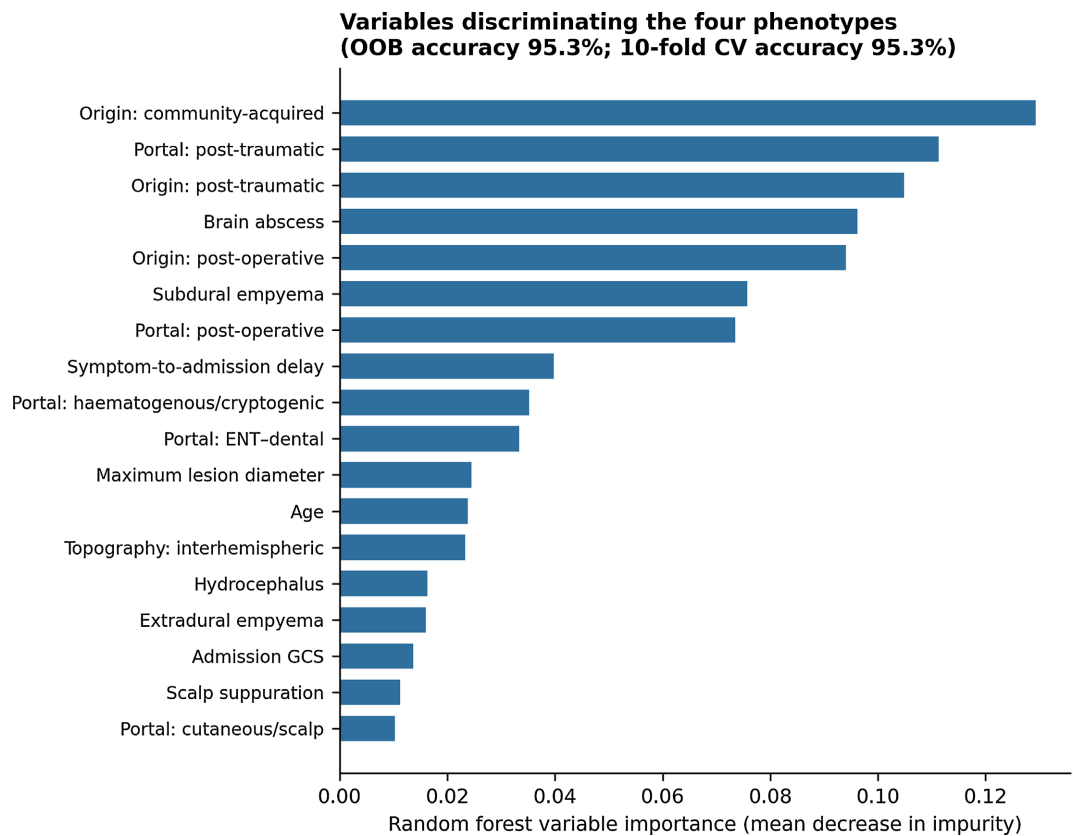


Figure S1. Variable importance in the restitution of the four-phenotype partition by random forest (2,000 trees). The origin, portal-of-entry and lesion-type variables dominate; the symptom-to-admission delay is the first of the quantitative variables, while age, the Glasgow score and lesion size contribute weakly. Only the 18 most contributory variables or categories are shown.