

Clinically Weighted Feature Fusion for Identifying Postpartum Psychological Risk Subgroups Among Breastfeeding Mothers

Eka Prasetyaningrum^{1,2,*}, Amiq Fahmi³, and Purwanto¹

¹ Master's Program in Informatics Engineering, Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang 50131, Indonesia; e-mail : p31202502711@mhs.dinus.ac.id; purwanto@dsn.dinus.ac.id

² Information Systems Program, Faculty of Computer Science, Universitas Darwan Ali, Sampit 74322, Indonesia; e-mail : eka.tya94@unda.ac.id

³ Information Systems Program, Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang 50131, Indonesia; e-mail : amiq.fahmi@dsn.dinus.ac.id

* Corresponding Author : Eka Prasetyaningrum 

Abstract: Breastfeeding experience is closely associated with postpartum psychological difficulties, including depression and anxiety, yet clinically relevant risk subgroups may remain insufficiently characterized. This study develops Clinically Weighted Feature Fusion (CWFF), a feature-weighting framework that integrates multiple breastfeeding and psychological domains by weighting features according to their Spearman correlations with EPDS and GAD-7 scores. The study analyzed 2,010 breastfeeding mothers in the United Kingdom, using a 70% training set ($n = 1,407$) and a held-out set ($n = 603$). CWFF weights and cluster centroids were estimated exclusively from the training data and fixed before application to the held-out data, thereby preventing observations from being used simultaneously for model formation and evaluation. The number of clusters was prespecified using the Silhouette Score, Davies-Bouldin Index, Calinski-Harabasz Index, and bootstrap Adjusted Rand Index stability across $k = 2-8$. The analysis identified $k = 2$, with a Silhouette Score of 0.213 and high clustering stability (ARI = 0.958). On the held-out data, CWFF achieved a Silhouette Score of 0.222, compared with 0.198 for equal weighting and 0.207 for PCA+K-Means, with a significant bootstrap difference from equal weighting (95% CI [+0.014, +0.033]). Independent validation using breastfeeding duration, which was used neither as a clustering feature nor to derive the CWFF weights, showed a significant difference between clusters (Kruskal-Wallis test, $H = 24.44$; $p < 0.001$), whereas maternal age did not differ significantly between clusters ($H = 1.52$; $p = 0.218$). However, the difference in clinical separation based on η^2_{EPDS} between CWFF and equal weighting was not statistically significant (95% CI [-0.005, +0.040]). Overall, the findings suggest that CWFF improves geometric separation of postpartum psychological risk subgroups while providing limited evidence of an additional clinical advantage over equal weighting.

Keywords: Breastfeeding; Clinically weighted feature fusion; Held-out validation; Postpartum anxiety; Postpartum depression; Psychological risk subgroups; Semi-supervised clustering; Spearman correlation.

Received: August, 10th 2026

Revised: September, 17th 2026

Accepted: September, 19th 2026

Published: October, 2nd 2026



Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) licenses (<https://creativecommons.org/licenses/by/4.0/>)

1. Introduction

Depression is a disorder typically characterized by persistent sadness, anhedonia, anxiety, and, in severe cases, a risk of self-harm [1]. Postpartum mothers are particularly susceptible to depression because of the substantial hormonal changes associated with the postpartum period [2]. However, awareness of maternal depression remains limited, making it an important public health concern because the condition often goes undiagnosed despite its potential consequences for both mothers and infants [3], [4]. Worldwide, approximately 17.2% of women are estimated to experience postpartum depression, with contributing factors including social conditions, economic status, and access to healthcare [5], [6].

Breastfeeding and lactation practices are also strongly associated with postpartum psychological well-being. Previous research has reported a reciprocal relationship between maternal mental health and breastfeeding experiences [7], [8]. Shorter breastfeeding duration and negative breastfeeding experiences have been associated with postpartum depression [9]. Common breastfeeding difficulties, such as breast pain and insufficient milk supply, may also contribute to guilt and feelings of maternal failure [10], [11]. In the UK, for example, a substantial gap has been reported between the intention to breastfeed (81%) and exclusive breastfeeding for six months (only 1%) [10]. These findings suggest that postpartum depression risk is shaped by interacting psychological, clinical, and breastfeeding-related factors, highlighting the need for analytical approaches capable of identifying complex patterns across these domains.

In response to this need, machine learning-based approaches for identifying postpartum depression risk are increasingly being developed because of their potential to model complex relationships among maternal risk factors [1], [12]–[14]. Most existing studies adopt supervised learning approaches based on binary diagnostic labels and use clinical variables, electronic medical records, or biomarkers that may not always be available in primary healthcare settings [1]. Other studies primarily focus on predicting depression status rather than identifying latent risk subgroups, even though mothers with similar EPDS scores may have different underlying risk profiles or contributing factors [15]. Feature engineering has also been widely applied to improve data representation; however, clustering is often performed directly on the engineered features without explicitly evaluating whether each clinical domain contributes meaningfully to the resulting cluster structure. Consequently, engineered features may be implicitly treated as equally informative, despite differences in the potential contribution of individual clinical domains to subgroup formation [16]. This limitation motivates the hypothesis that clinically informed feature representation and weighting may provide a more meaningful basis for identifying postpartum psychological risk subgroups than equal treatment of heterogeneous features.

This study builds on an earlier unsupervised approach developed by the authors' team using the same dataset [17]. In that study, subgroup identification was performed using K-Means on a standardized feature space comprising 30 raw variables. Principal Component Analysis (PCA) was used to examine variance distribution and facilitate component interpretation. The number of clusters was set to $k = 4$ based on clinical interpretability; however, no predefined selection criterion or formal decision rule was established before the analysis. To address this limitation, the present study develops a clinically informed feature representation framework based on a validation-before-weighting paradigm. Rather than directly applying engineered features to clustering, the proposed approach first constructs a feature representation across five clinically meaningful domains: breastfeeding-related stress, psychological burden, breastfeeding challenges, mental health history, and cross-domain interactions, grounded in clinical evidence [18], [19]. The resulting representation is then empirically evaluated through ablation analyses, including leave-one-domain-out and leave-one-interaction-term-out procedures, bootstrap testing for robustness, and comparative evaluation across multiple clustering algorithms. This process enables the contribution of individual domains to be examined before feature weighting is applied.

Building on this validated feature representation, the study introduces Clinically Weighted Feature Fusion (CWFF), a feature-weighting mechanism that uses training-data-based associations with EPDS and GAD-7 scores to construct a representation more closely aligned with the psychological characteristics of postpartum mothers. Rather than assuming that all domains contribute equally, CWFF assigns weights based on the empirical strength of each feature's association with these clinical measures. The proposed framework is therefore guided by the second hypothesis that clinically informed feature weighting can improve the separation of postpartum psychological risk subgroups while maintaining an evaluation process that prevents information leakage from held-out data.

Specifically, the present study extends the previous work through four methodological enhancements. First, it introduces the CWFF framework, which explicitly assigns feature weights according to their clinical relevance to EPDS and GAD-7 outcomes. Second, the number of clusters (k) is determined using a predefined selection rule based on sensitivity and statistical stability criteria, replacing the post hoc selection of $k = 4$ used in the previous study. Third, a train–held-out evaluation scheme separates feature-weight estimation and cluster-centroid construction from the subsequent evaluation of their effectiveness. Fourth, an

independent validation strategy uses variables that are entirely excluded from model development. Together, these enhancements provide a more controlled framework for evaluating clinically informed subgroup identification while distinguishing the present methodology from the preceding study.

The primary methodological contribution of this study therefore lies not in developing a new clustering algorithm, but in integrating clinically informed feature construction, domain-level validation, robustness assessment, and feature weighting within a unified workflow. The five-domain feature representation is first evaluated empirically, after which feature weights are estimated exclusively from the training data based on the strength of their associations with EPDS and GAD-7 scores. The resulting weights are then fixed before being applied to the held-out data, preventing information from the evaluation set from influencing model development. This design allows the proposed framework to place greater emphasis on clinically relevant features while retaining a clear separation between model development and evaluation. On the held-out data, the approach demonstrates improved cluster separation compared with equal weighting (Section 4.4). Accordingly, the novelty of this study lies in the design and systematic evaluation of a clinically informed feature representation framework rather than in the introduction of a new clustering algorithm. The proposed framework may provide an initial basis for exploring postpartum psychological risk subgroups with greater clinical relevance. However, the findings should not yet be interpreted as sufficient evidence for clinical implementation or screening, as further validation using independent datasets, longitudinal designs, and more diverse populations remains necessary.

2. Related Work

2.1. Machine Learning for Postpartum Depression Prediction

Most published machine learning models frame postpartum depression detection as a supervised binary classification problem. Lin et al. [3] developed a simple indicator model for practical clinical use, while Fazraningtyas et al. [4] applied ensemble learning to improve prediction accuracy on imbalanced maternal health data. Multimodal and generative approaches have also been explored. García-Méndez and De Arriba-Pérez [1] applied generative AI for real-time detection and explanation, whereas Zhang et al. [20] combined text, audio, and video signals for broader depression assessment. Gulmohamed et al. [18] proposed a privacy-preserving pipeline for early prediction using multidimensional maternal health data, while Garbaza et al. [14] modeled perinatal depression risk from psychosocial and sleep-related factors. Across this literature, the systematic review by Fazraningtyas et al. [21] notes that many models still rely on numerous clinical or biomarker-based variables, which may limit their transferability to healthcare facilities without electronic medical records.

A recurring limitation of this supervised paradigm is that binary depressed/not-depressed labels may mask clinically meaningful heterogeneity within the population. Orchard et al. [12] addressed this issue in the broader mental health context by applying unsupervised clustering to service-use patterns in Ontario, Canada. Their findings showed that latent subgroups with distinct service-utilization profiles can persist even within a single diagnostic category. This suggests that subgroup-oriented analysis may provide complementary information to binary prediction by capturing differences in psychological and behavioral profiles that are not represented by a single diagnostic label.

2.2. Feature Engineering and Clustering on Health Survey Data

Clustering algorithms such as K-Means are sensitive to how the input feature space is constructed, and studies beyond maternal health illustrate the importance of domain-grounded feature representation. Nedyalkova and Madurga [16] combined K-Means with combinatorial feature selection for Type 2 diabetes phenotyping, while Galankin et al. [22] cautioned against the uncritical use of PCA when clustering behavioral and psychiatric symptom data. Their work suggests that dimensionality reduction may obscure clinically meaningful symptom patterns when applied without sufficient domain grounding. Beyond feature representation, standardization is also important for distance-based clustering. Usman and Mohamad [23] showed that differences in variable scales can distort Euclidean distance calculations, meaning that inadequate feature scaling may degrade clustering quality, particularly for K-Means.

In the maternal health context, Mathkor [24] examined the relationships among maternal work stress, breastfeeding experience, mental health, and infant attachment using a combination of machine learning and statistical modeling. The study emphasized the importance of considering these factors as an interacting system rather than as independent predictors. Meanwhile, Rowles et al. [19] provided a clinical foundation for considering breastfeeding difficulties as multidimensional factors affecting maternal mental health through physical, social, and psychological mechanisms, using the same cohort as the present study.

Despite these advances, an important methodological gap remains. Existing studies generally focus either on predicting postpartum depression outcomes or on applying clustering to a predefined feature space, with limited attention to whether heterogeneous clinical domains contribute differently to the resulting subgroup structure. In particular, there is limited evidence of a systematic framework that combines domain-based feature construction, empirical assessment of domain contributions, and clinically informed feature weighting while maintaining strict separation between model development and evaluation. This gap motivates the development of Clinically Weighted Feature Fusion (CWFF), which integrates multiple clinical domains into a weighted feature representation and evaluates their contributions through ablation analysis, bootstrap assessment, and statistical testing.

3. Proposed Method

This section presents the methodology used to develop and evaluate the Clinically Weighted Feature Fusion (CWFF) framework. The overall design emphasizes clinically informed feature representation while maintaining a clear separation between model development and evaluation. The methodology comprises dataset preparation, clinical domain-based feature engineering, train/held-out splitting, cluster-number selection, CWFF weighting, and subsequent validation and robustness analyses.

3.1. Dataset

This study uses the open-access dataset provided by Braithwaite et al. [25], comprising 2,010 postpartum mothers in the United Kingdom who were breastfeeding their first child. The dataset was collected under ethical approval from Manchester Metropolitan University (REF 58254). It contains 140 variables covering lactation-related pressure, psychosocial impact, breastfeeding challenges, mental health history, and EPDS/GAD-7 measures.

Because the objective of this study is to identify clinically meaningful subgroups rather than to use all available survey variables indiscriminately, only variables relevant to the predefined clustering representations were selected. The selection criteria and rationale for the two feature representations used in this study are described in Section 3.4. The EPDS and GAD-7 measures were retained for clinical association analysis and CWFF weight estimation, where applicable, but were excluded from the clustering feature space to avoid directly defining the clusters through the outcomes used for evaluation.

3.2. Research Workflow

The research workflow consists of seven stages, as illustrated in Figure 1. First, the dataset is loaded and preprocessed. Second, two complementary feature representations are constructed in parallel: a raw-variable representation with mental health history variables (Approach A) and a clinically informed, domain-based engineered representation (Approach B). Third, the data are divided into training (70%) and held-out (30%) sets using stratification based on $EPDS \geq 10$ risk status. Fourth, feature standardization is fitted exclusively on the training data and subsequently applied to the held-out data without refitting. Fifth, the number of clusters is determined using a prespecified selection rule based only on the training data. Sixth, K-Means and CWFF are developed using the training data, with the CWFF weights and cluster centroids fixed before application to the held-out data. Finally, the resulting cluster structures are evaluated using clustering-quality measures, clinical association analyses, independent validation variables, bootstrap testing, ablation analysis, and comparisons with alternative clustering approaches.

This workflow is designed to maintain a clear separation between model development and evaluation. In particular, $EPDS_total$ and $GAD-7_total$ are excluded from the clustering feature space and are not used as direct clustering inputs. This prevents the resulting clusters from being defined directly by the clinical measures subsequently used to characterize them.

The use of a held-out dataset also allows the clustering representation and CWFF weighting scheme to be evaluated on observations that were not involved in their construction.

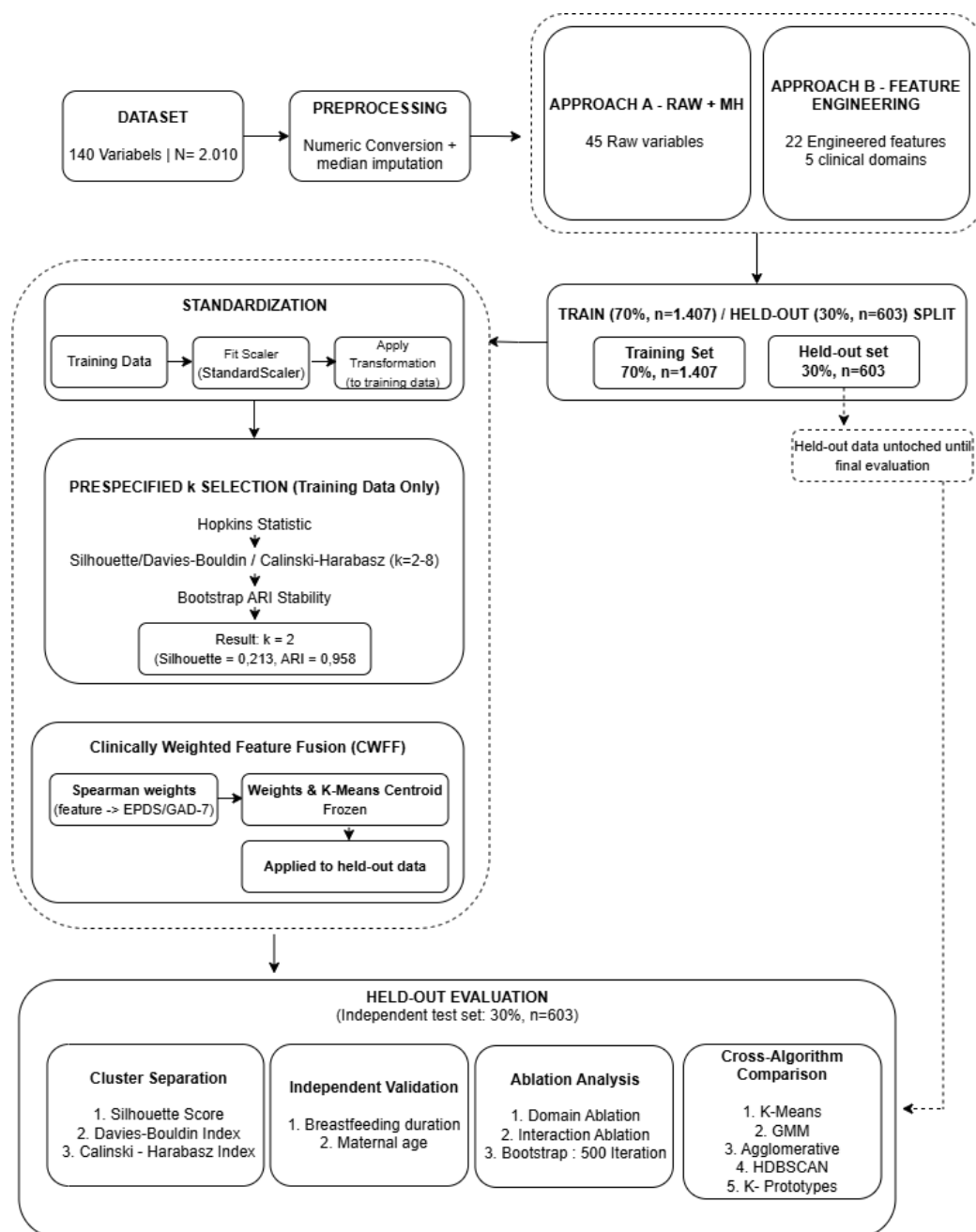


Figure 1. Research Workflow.

3.3. Clinical Domain-Based Feature Engineering

The main methodological contribution of this study is the transformation of raw variables into 22 features organized into five clinically meaningful domains: Pressure, Psychological Burden, Breastfeeding Challenges, Mental Health History, and cross-domain Interaction Features. This domain-based representation is intended to summarize related clinical factors while maintaining their interpretability for postpartum psychological risk analysis. Five interaction features are additionally constructed through multiplication to capture potential cross-domain relationships, particularly the interaction between psychological burden and pressure. All features are constructed through addition and multiplication of Likert-scale and binary variables and are subsequently used as the clustering input without combining them with the raw variables. A summary of the feature engineering framework is presented in Table 1.

Table 1. Clinical domain-based feature engineering framework (Approach B)

Clinical Domain	Number of Features	Formula / Basis	Clinical Rationale
Pressure Score	5	Sum of prenatal/postnatal pressure items; pressure change; healthcare-provider and social-source subtotals	Represents pressure accumulated across different sources rather than treating individual sources independently
Psychological Burden	5	Sum of internal burden, social impact, functional impact, and anxiety sub-indices	Represents psychological burden across multiple related dimensions
Breastfeeding Challenges	3	Sum of nine binary challenge items; physical and milk-supply subtotals	Represents breastfeeding-related difficulties while distinguishing physical burden from milk-supply disruption
Mental Health History	4	Presence flag; severity score = $\sum w_i \cdot MH_i$, where $w_i \in \{1, 2\}$	Represents the presence and severity of reported mental health history
Interaction Features	5	Five pairwise products across Domains 1–4, including psychological burden $\times$ pressure	Represents potential cross-domain interactions that may not be captured by individual domains alone

Each domain is formed through simple arithmetic operations to produce composite features that provide a more structured representation of the underlying survey variables. The Healthcare and Social Pressure domain measures prenatal and postnatal pressure by summing the respective items. The change between these two conditions is then calculated, together with separate pressure measures based on healthcare-provider and social sources.

$$FE_{prenatal} = \sum_{i=1}^8 pre_i \quad (1)$$

$$FE_{postnatal} = \sum_{i=1}^9 post_i \quad (2)$$

$$FE_{pressure_delta} = FE_{postnatal} - FE_{prenatal} \quad (3)$$

The Psychological Burden domain combines internal burden, social impact, functional impact, and anxiety into a single composite score to represent related dimensions of postpartum psychological experience.

$$FE_{psycho} = internal_burden + social_impact + functional_impact + anxiety_impact \quad (4)$$

The Breastfeeding Challenges domain is calculated from the sum of nine binary breastfeeding difficulty indicators, with a higher score indicating more challenges experienced. The Mental Health History domain combines mental health history with a weight of 1 for suspected conditions and a weight of 2 for diagnosed conditions.

$$FE_{MH} = \sum_j w_j \cdot MH_j, \quad w_j \in \{1, 2\} \quad (5)$$

To capture cross-domain relationships that may not be represented by individual domains alone, Interaction Features are formed through cross-domain multiplication. The primary interaction represents the relationship between psychological burden and postnatal pressure:

$$FE_{psycho \times pressure} = FE_{psycho} \times FE_{postnatal} \quad (6)$$

Beyond this interaction, several other cross-domain combinations are constructed to represent potential relationships among the clinical factors contributing to postpartum psychological risk.

For the Psychological Burden domain, which is constructed as a reflective construct, internal consistency was assessed using Cronbach's alpha on the training data. The results indicated that Internal Burden ($\alpha = 0.904$) and Anxiety Impact ($\alpha = 0.828$) showed high internal consistency, while Functional Impact ($\alpha = 0.615$) and Social Impact ($\alpha = 0.714$) showed adequate consistency. All corrected item-total correlations were above 0.3, indicating that each item was consistent with its respective block. Meanwhile, Pressure, Breastfeeding Challenges, and Mental Health History were treated as formative indices; therefore, Cronbach's alpha was not applied to these domains. Their contribution to the cluster structure was instead evaluated through the ablation analysis described in Section 3.8.

3.4. Standardization, Raw+MH Baseline, and Train/Held-Out Scheme

To assess whether the clinically informed feature representation provides additional value beyond the use of raw variables, two complementary representations are considered. Approach A serves as the raw-variable baseline, whereas Approach B uses the 22 engineered features described in Section 3.3. The baseline was designed to provide a fair comparison by ensuring that mental health history information is available in both approaches. In the previous study, Approach A did not include these variables, whereas the proposed representation incorporated them; therefore, differences between the approaches could not be attributed solely to the feature construction strategy [17].

Approach A (raw+MH) consists of 45 raw variables selected from the 140 variables available in the dataset: 17 pressure variables, 11 psychological impact variables, 9 breastfeeding challenge variables, and 8 mental health history variables. The free-text variable MH_describe was excluded. Missing values in the mental health history variables were assigned a value of 0, indicating that the corresponding option was not selected. EPDS/GAD-7 items, demographic variables, and administrative variables were excluded from both approaches so that the clustering representation remained focused on the five clinical domains.

The 2,010 observations were divided into training and held-out sets using a 70:30 split, resulting in 1,407 training observations and 603 held-out observations. Stratification based on EPDS ≥ 10 risk status produced nearly identical proportions between the two sets (48.3% in the training set and 48.4% in the held-out set). Standardization was performed using a StandardScaler fitted exclusively on the training data and subsequently applied to the held-out data without refitting. The determination of the number of clusters and the CWFF feature weights likewise used only the training data. The held-out data were reserved for the final evaluation of the resulting cluster representation.

This design separates model development from evaluation and ensures that information from the held-out observations does not influence feature scaling, cluster-number selection, feature weighting, or cluster-centroid construction. Accordingly, the performance reported on the held-out data reflects the application of a representation and clustering model developed without using those observations.

3.5. Cluster Feasibility Test and k Selection

Before determining the number of clusters k , the Hopkins Statistic was calculated on the training data to assess whether the data exhibited a natural tendency to form clusters. A value of 0.5 indicates a pattern close to random, whereas a value near 1 indicates a clearer tendency toward cluster structure.

Unlike the previous study, which determined k based on clinical considerations after observing the clustering results, the present study establishes the k -selection rule before examining the clinical characteristics of the resulting clusters. This approach is intended to reduce the potential for post hoc selection of the number of clusters based on their subsequent clinical interpretation.

The selection process consists of three stages: (1) testing $k = 2$ through $k = 8$ using the Silhouette Score, Davies-Bouldin Index, and Calinski-Harabasz Index; (2) selecting candidate values whose Silhouette Score falls within 5% of the best observed value; and (3) selecting the candidate with the highest stability based on the mean Adjusted Rand Index (ARI) across 100 bootstrap samples of the training data. This procedure allows the number of clusters to be determined using predefined clustering-quality and stability criteria before the clinical characteristics of the clusters are examined.

3.6. Clinically Weighted Feature Fusion (CWFF)

The CWFF framework is designed to address the possibility that the engineered features do not contribute equally to the identification of postpartum psychological risk subgroups. Rather than assigning the same weight to all features, CWFF uses their empirical associations with EPDS and GAD-7 scores to reflect differences in clinical relevance. The feature weights are estimated exclusively from the training data and fixed before being applied to the held-out data. For each feature f , the initial weight is calculated as the average absolute Spearman correlation with the total EPDS and GAD-7 scores in the training data:

$$w_f = \frac{|\rho(f, EPDS)| + |\rho(f, GAD - 7)|}{2} \quad (7)$$

The resulting weight is then normalized by the mean weight across all features:

$$w'_f = \frac{w_f}{\bar{w}} \quad (8)$$

where $\bar{w}$ denotes the mean weight across all features. This normalization preserves the relative differences among features while maintaining the overall scale of the weighted representation.

The training data are first standardized into z-scores using a scaler fitted exclusively on the training data and are subsequently multiplied by the corresponding feature weights. Standardization is performed before weighting so that differences in the original feature scales do not obscure the effect of the CWFF weights. If weighting were applied before standardization, the subsequent scaling process could reduce or eliminate the intended differences in feature contribution.

The K-Means centroids for CWFF are trained using only the weighted training data. For the held-out data, the same standardization parameters and fixed CWFF weights are applied, and observations are assigned to the previously established cluster centroids without retraining the model. This procedure ensures that the information used to determine feature weights and cluster centroids is derived only from the training data.

Because CWFF uses clinical information from EPDS and GAD-7 to determine feature weights, the approach is treated as semi-supervised weighting. Approach B, which uses the same engineered features without clinically informed weighting, serves as the equal-weight unsupervised baseline. This distinction allows the evaluation to examine whether incorporating clinically informed feature contributions provides additional value beyond the underlying domain-based feature representation itself.

3.7. Statistical Validation

Statistical analyses were performed to characterize differences between the identified clusters and assess their associations with clinically relevant variables. Data normality was assessed using the Shapiro–Wilk test, and homogeneity of variance was evaluated using Levene's test. Because most variables were not normally distributed, differences between clusters were assessed using the Kruskal–Wallis test, followed by the Dunn–Bonferroni test when significant. Effect size was measured using eta-squared (η^2)

$$\eta^2 = \frac{H}{N - 1} \quad (9)$$

where H is the Kruskal–Wallis test statistic and N is the total number of observations.

EPDS and GAD-7 were used to assess clinical association but were not considered independent validation variables because they were used to estimate the CWFF weights. Independent validation was instead performed using breastfeeding duration and maternal age, neither of which was used as a clustering feature or to derive the CWFF weights. Differences in Silhouette Score and η^2 between CWFF and equal weighting were evaluated using 1,000 bootstrap iterations on the held-out data, with 95% confidence intervals estimated using the percentile method:

$$\Delta M = M_{CWFF} - M_{Equal-weight} \quad (10)$$

This analysis provides an estimate of the uncertainty surrounding the observed differences rather than relying solely on point estimates from the held-out sample.

3.8. Ablation Design

Ablation analysis was performed to examine the contribution of individual clinical domains and interaction features to the resulting cluster structure. A leave-one-domain-out procedure was applied by removing each of the five domains in turn. The CWFF weights and K-Means centroids were retrained using the training data and subsequently evaluated on the same held-out data as the full model.

The same procedure was applied to each interaction feature using a leave-one-interaction-term-out design. Changes in Silhouette Score relative to the full 22-feature CWFF model were assessed using 500 bootstrap iterations on the held-out data. The trained model was kept fixed during resampling, with only the held-out observations resampled.

3.9. Comparison Algorithms and Cross-Algorithm Significance

To examine whether the observed clustering behavior depends on the clustering algorithm, the feature representations were additionally evaluated using Gaussian Mixture Model (GMM), Agglomerative Clustering, and HDBSCAN. For HDBSCAN, `min_cluster_size` was evaluated at {15, 25, 40, 60, 80, 100, 150}, while `min_samples` was tested at [5, 10, default].

Two additional comparisons were conducted. PCA+K-Means was applied to Approach A using 22 principal components, matching the dimensionality of the CWFF representation and explaining 95.93% of the total variance. K-Prototypes was also evaluated because several engineered features, including `FE_prior_mh_any` and `FE_physical_challenge`, are binary or ordinal rather than fully continuous. For these comparisons, model fitting and transformation followed the same train/held-out principle used in the primary analysis.

4. Results and Discussion

4.1. Cluster Feasibility and k Selection

The Hopkins statistic calculated on the training data was 0.791, indicating a clear tendency toward cluster structure rather than random dispersion. This result supports the suitability of clustering analysis and provides an initial basis for proceeding with K-Means. The sensitivity analysis for $k = 2$ –8 on the training data is presented in Table 2.

Table 2. Sensitivity of $k = 2$ –8 on the Training Data ($n = 1,407$)

k	Silhouette	Davies–Bouldin	Calinski–Harabasz
2	0.213	1.708	446.6
3	0.201	1.848	337.5
4	0.153	1.790	296.5
5	0.142	1.780	258.9
6	0.142	1.761	235.6
7	0.146	1.752	219.5
8	0.143	1.822	204.0

As shown in Table 2, $k = 2$ provides the strongest overall clustering structure across the three evaluation metrics and is the only candidate that falls within 5% of the best Silhouette Score. The corresponding trends in the Silhouette Score and Davies–Bouldin Index are illustrated in Figure 2, where the dashed vertical line indicates the selected $k = 2$. The bootstrap stability analysis further supported this choice, yielding a mean Adjusted Rand Index (ARI) of 0.958 (SD = 0.032) for $k = 2$ indicating high consistency under resampling. Accordingly, based on the pre-specified selection rule, $k = 2$ was retained as the final number of clusters.

The same $k = 2$ solution was also obtained in two complete analysis runs using the same random state, providing additional evidence that the selected cluster structure was reproducible under the specified analysis settings. Taken together, the feasibility, sensitivity, and stability analyses provide a stable and prespecified basis for the subsequent held-out evaluation.

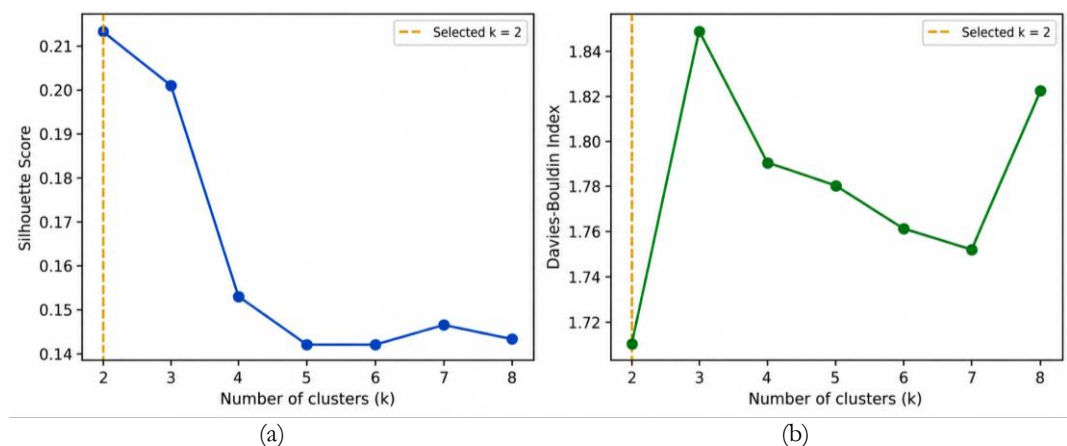


Figure 2. Sensitivity of cluster number $k = 2$ –8 on the training data: (a) Silhouette Score and (b) Davies–Bouldin Index. The dashed vertical line indicates the selected $k = 2$.

4.2. Clinically Weighted Feature Fusion Weights

The CWFF weights obtained from the training data are presented in Table 3. The weighting pattern shows that features from the psychological burden and mental health history domains generally receive greater weights, whereas features representing physical breastfeeding challenges receive comparatively lower weights.

Table 3. CWFF weights assigned to the 22 engineered features based on their associations with EPDS and GAD-7 scores in the training data.

Rank	Feature	CWFF Weight	Rank	Feature	CWFF Weight
1	FE_mh_severity	1.839	12	FE_total_postnatal_pressure	0.9585
2	FE_prior_mh_any	1.508	13	FE_hcp_x_guilt	0.9455
3	FE_mh_x_challenges	1.421	14	FE_total_prenatal_pressure	0.9170
4	FE_total_psycho_burden	1.375	15	FE_anxiety_impact	0.8461
5	FE_psycho_x_pressure	1.348	16	FE_social_pressure	0.8248
6	FE_internal_burden	1.232	17	FE_hcp_pressure	0.7727
7	FE_prior_anxiety	1.195	18	FE_pressure_x_challenges	0.7151
8	FE_functional_impact	1.163	19	FE_pressure_delta	0.4446
9	FE_social_impact	1.154	20	FE_total_challenges	0.4303
10	FE_burden_per_challenge	1.0799	21	FE_physical_challenge	0.4023
11	FE_prior_depression	1.0719	22	FE_supply_challenge	0.3383

As shown in Table 3, the highest-weighted features include FE_mh_severity (1.839), FE_prior_mh_any (1.508), FE_mh_x_challenges (1.421), and FE_total_psycho_burden (1.375). In contrast, the lowest weights are assigned to FE_physical_challenge (0.4023) and FE_supply_challenge (0.3383). Overall, this pattern places greater emphasis on psychological and mental health history features than on physical breastfeeding challenges within the CWFF representation.

This weighting pattern is also consistent with the principal component analysis results, in which guilt and threats to maternal identity account for more of the observed data variation than physical breastfeeding challenges. Thus, the CWFF representation appears to emphasize psychological and prior mental health characteristics that are more closely aligned with the psychological outcomes used to derive the weights, while retaining physical breastfeeding challenges as part of the broader feature representation.

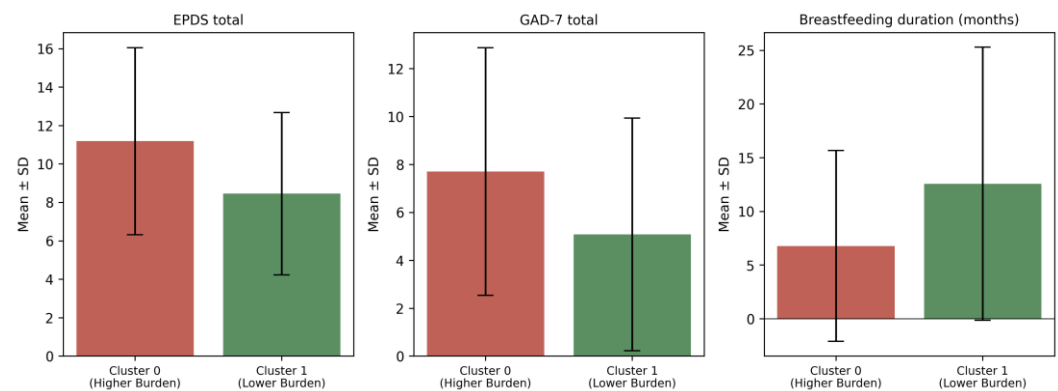
4.3. Cluster Profile ($k = 2$)

At $k = 2$, the 603-sample held-out dataset was divided into two clusters: Cluster 0 comprised 338 samples (56.1%), whereas Cluster 1 comprised 265 samples (43.9%). The outcome profiles of the two clusters are summarized in Table 4.

Table 4. Outcome Profile by Cluster on the Held-Out Data ($n = 603$)

Variable	Cluster 0 (mean $\pm$ SD)	Cluster 1 (mean $\pm$ SD)	η^2	p
EPDS_total	11.18 $\pm$ 4.86	8.45 $\pm$ 4.22	0.078	<0.001
GAD7_total	7.70 $\pm$ 5.17	5.08 $\pm$ 4.86	0.084	<0.001
Breastfeeding duration (months)	6.77 $\pm$ 8.88	12.57 $\pm$ 12.72	0.048	<0.001
Maternal age (years)	34.88 $\pm$ 7.76	34.89 $\pm$ 5.18	0.001	0.218

As shown in Table 4, Cluster 0 exhibits higher EPDS and GAD-7 scores and a shorter breastfeeding duration than Cluster 1. The three indicators show a consistent pattern, with greater psychological symptom burden accompanied by a shorter breastfeeding duration in Cluster 0. In contrast, maternal age does not differ significantly between the clusters ($p = 0.218$), suggesting that the observed cluster differentiation is not primarily associated with age. The corresponding profiles of EPDS, GAD-7, and breastfeeding duration are illustrated in Figure 3.

**Figure 3.** EPDS, GAD-7, and breastfeeding duration profile by cluster (mean $\pm$ SD) on the held-out data ($n = 603$).**Table 5.** Top Ten Engineered Features Differentiating the Two Clusters Based on η^2 , Held-Out Data

Feature	Cluster 0 (mean)	Cluster 1 (mean)	η^2
FE_pscho_x_pressure	1034.55	314.83	0.676
FE_total_pscho_burden	73.07	30.17	0.644
FE_internal_burden	24.74	8.58	0.546
FE_hcp_x_guilt	83.23	21.87	0.489
FE_anxiety_impact	16.16	7.23	0.449
FE_functional_impact	18.74	9.07	0.383
FE_burden_per_challenge	18.56	8.44	0.371
FE_social_impact	13.43	5.29	0.289
FE_total_postnatal_pressure	14.06	10.45	0.259
FE_total_prenatal_pressure	12.09	9.18	0.240

To further characterize the factors underlying the observed cluster differences, the engineered features were examined using η^2 . As shown in Table 5, several engineered features exhibit larger between-cluster effect sizes than the direct clinical outcomes in Table 4. In particular, FE_pscho_x_pressure, FE_total_pscho_burden, and FE_internal_burden show the largest η^2 values, indicating substantial differences between the two cluster profiles. The pattern also highlights the role of features that combine psychological burden, pressure, and internal experiences in distinguishing the clusters. These findings suggest that the engineered representation provides additional detail for characterizing cluster heterogeneity beyond the direct screening scores. However, the observed between-cluster differences should be interpreted as differences in feature profiles rather than evidence of greater clinical validity of

individual engineered features. Based on the observed characteristics, the clusters are therefore assigned neutral descriptive labels: Cluster 0 is referred to as “Higher Psychological Burden,” reflecting its higher EPDS and GAD-7 scores and shorter breastfeeding duration, whereas Cluster 1 is referred to as “Lower Psychological Burden.” These labels are intended solely to facilitate interpretation and do not represent diagnostic or clinical categories.

4.4. Held-Out Evaluation and Method Comparison

The clustering performance of the different approaches on the held-out data is summarized in Table 6. CWFF achieved a Silhouette Score of 0.222, compared with 0.198 for the equal-weight Approach B, corresponding to a mean difference of +0.024. Based on 1,000 bootstrap iterations, the 95% confidence interval for this difference was [+0.014, +0.033], which excludes zero and indicates that the observed improvement was consistent across the bootstrap resamples. CWFF also achieved a higher Silhouette Score than PCA + K-Means (0.207).

Table 6. Clustering Performance of the Compared Methods on the Held-Out Data ($n = 603$)

Method	Silhouette	Davies–Bouldin	Calinski–Harabasz
Approach B (equal-weight)	0.198	—	—
PCA + K-Means (22 components)	0.207	1.834	165.3
CWFF (Approach C)	0.222	1.604	201.5

The comparison in Table 6 indicates that CWFF provides improved geometric separation relative to both equal weighting and PCA-based dimensionality reduction under the held-out evaluation. The improvement over equal weighting is particularly relevant because both approaches use the same engineered feature representation, with the principal difference being the clinically informed weighting applied by CWFF. Thus, the observed difference provides evidence that feature weighting contributes to the improvement in cluster separation, rather than the improvement being attributable solely to feature engineering.

For an additional comparison with mixed-type clustering, CWFF-K-Means and K-Prototypes produced an ARI of 0.514 on the training data. This moderate agreement indicates that the two methods identify partially overlapping cluster structures despite relying on different assumptions regarding feature types. At the same time, the ARI being below 1 indicates that a proportion of observations receive different cluster assignments across the two methods. This comparison therefore supports the presence of related, but not identical, cluster structures across the two clustering approaches.

Overall, the held-out results suggest that the principal benefit of CWFF in this analysis is improved geometric separation of the identified subgroups. This result provides a basis for the subsequent analyses examining whether the improved separation is also reflected in clinically relevant cluster differences and whether the observed pattern is maintained across alternative representations and clustering algorithms.

4.5. Clinical Validation: Association versus Independent Validation

The clinical and independent validation results on the held-out data are summarized in Table 7. EPDS and GAD-7 were treated as clinical association measures because, although they were not used as clustering features, they were used to derive the CWFF weights, whereas breastfeeding duration and maternal age were used neither as clustering features nor to derive the CWFF weights and therefore provide independent validation measures.

Table 7. Kruskal–Wallis Test Results for Clinical Association and Independent Validation on the Held-Out Data ($n = 603$)

Category	Variable	H	η^2	p
Clinical association	EPDS_total	47.74	0.078	<0.001
Clinical association	GAD7_total	51.53	0.084	<0.001
Independent validation	Breastfeeding duration	24.44	0.048	<0.001
Independent validation	Maternal age	1.52	0.001	0.218

As shown in Table 7, both EPDS and GAD-7 differ significantly between the two clusters, consistent with the clinical associations incorporated into the CWFF weighting process. Breastfeeding duration, which was used neither as a clustering feature nor to derive the CWFF weights, also differs significantly between the clusters on the held-out data. Because this variable was excluded from the clustering process, its association with the resulting clusters provides an independent, non-circular validation of the identified subgroup structure. In contrast, maternal age does not differ significantly between clusters ($p = 0.218$), indicating that the observed subgroup differentiation is not similarly reflected in this demographic characteristic.

The comparison between CWFF and equal weighting provides an additional perspective on the extent to which clinically informed weighting improves cluster separation. Although CWFF produced a difference of +0.016 in η^2 for EPDS relative to Approach B, the corresponding 95% CI of [-0.005, +0.040] includes zero. Thus, the held-out analysis does not provide sufficient evidence that CWFF achieves significantly greater EPDS separation than equal weighting.

This result differs from the earlier analysis, in which a significant CWFF advantage was observed for both η_{EPDS}^2 and η_{GAD7}^2 . Under the present held-out evaluation, which was designed to reduce circularity between model development and evaluation, that advantage is no longer statistically supported. Overall, the findings distinguish two aspects of CWFF performance: the method provides clearer geometric separation of the clusters, whereas the evidence for an additional improvement in clinical indicator separation over equal weighting remains limited. This distinction is important when interpreting the contribution of clinically informed feature weighting and avoiding overinterpretation of the clustering results.

4.6. Ablation Results: Domain and Interaction-Term Contributions

Each ablation variant was constructed independently. For each variant, preprocessing, CWFF weighting, and K-Means clustering were retrained using only the training data and subsequently evaluated on the same held-out data as the main model ($k = 2$). The change in Silhouette Score was used to assess changes in geometric separation quality and, by itself, does not indicate the clinical superiority or inferiority of a particular feature configuration.

4.6.1. Domain Ablation

The domain-level ablation results are presented in Table 8.

Table 8. Domain Ablation Results on the Held-Out Data ($k = 2$). Δ Silhouette was computed from unrounded Silhouette values; last-digit differences from the rounded values shown are due to rounding.

Variant	Domain Removed	Silhouette	Δ Silhouette	Bootstrap 95% CI	Significant
Full CWFF (22 features)	—	0.2223	—	—	—
Without Pressure	Pressure	0.3035	+0.0812	[0.055, 0.106]	Yes
Without Psychological Burden	Psychological Burden	0.3417	+0.1194	[0.093, 0.144]	Yes
Without Breastfeeding Challenges	Breastfeeding Challenges	0.2268	+0.0046	[0.004, 0.006]	Yes
Without MH History	Mental Health History	0.2947	+0.0724	[0.060, 0.085]	Yes
Without Interaction Features	Interactions	0.2723	+0.0500	[0.025, 0.075]	Yes

As shown in Table 8, removing each domain increases the Silhouette Score relative to the full CWFF configuration, with the largest increase observed after removing the Psychological Burden domain (+0.1194; bootstrap 95% CI [0.093, 0.144]). Removing the Pressure, Mental Health History, and Interaction domains also increases geometric separation, although to different extents.

This pattern reflects the distinction between the two objectives considered in this study. CWFF assigns weights according to the relationships between features and the clinical indicators EPDS and GAD-7, whereas the Silhouette Score evaluates geometric separation within

the feature space. Consequently, a feature or domain receiving a relatively high CWFF weight because of its clinical association does not necessarily improve geometric separation. An increase in Silhouette Score after removing a domain should therefore not be interpreted as evidence that the corresponding domain lacks clinical relevance.

The full 22-feature configuration is retained as the main model because it preserves all domains specified in the conceptual framework and maintains the broader clinical coverage of the proposed representation. Accordingly, the ablation results are interpreted primarily as a sensitivity analysis, highlighting a trade-off between geometric separation and the coverage of clinically relevant information. Overall, the findings suggest that maximizing the Silhouette Score alone would favor a more reduced representation, whereas the full CWFF configuration preserves the multidimensional structure intended by the proposed framework.

4.6.2. Interaction-Term Ablation

The contribution of the individual interaction terms was further examined by removing each term separately. The results are presented in Table 9.

Table 9. Interaction-Term Ablation Results on the Held-Out Data ($k = 2$). Δ Silhouette was computed from unrounded Silhouette values; last-digit differences from the rounded values shown are due to rounding.

Interaction Removed	Silhouette	Δ Silhouette	Bootstrap 95% CI	Significant
FE_psycho_x_pressure	0.2780	+0.0558	[0.031, 0.079]	Yes
FE_hcp_x_guilt	0.2728	+0.0505	[0.025, 0.073]	Yes
FE_burden_per_challenge	0.2763	+0.0540	[0.029, 0.079]	Yes
FE_mh_x_challenges	0.2350	+0.0127	[0.007, 0.018]	Yes
FE_pressure_x_challenges	0.2240	+0.0017	[-0.0002, 0.0030]	No

As shown in Table 9, removing four of the five interaction terms produces a statistically significant increase in the Silhouette Score relative to the full CWFF configuration. The largest increase is observed after removing FE_psycho_x_pressure, with a reported Δ Silhouette of +0.0558 and a bootstrap 95% CI of [0.031, 0.079]. Similar increases are observed for FE_burden_per_challenge and FE_hcp_x_guilt, whereas removing FE_mh_x_challenges produces a comparatively smaller change.

In contrast, removing FE_pressure_x_challenges produces only a small change in geometric separation, and its 95% CI [-0.0002, 0.0030] includes zero. Thus, its contribution to the observed Silhouette Score is not statistically distinguishable from zero under the present ablation analysis. This term may therefore warrant further examination in future model simplification studies, although any decision to remove it should consider not only geometric separation but also clinical relevance and the stability of the resulting cluster profiles.

Taken together, the domain- and interaction-level ablations indicate that several clinically motivated components reduce geometric separation when evaluated solely by the Silhouette Score. Rather than contradicting the CWFF rationale, this finding highlights the difference between optimizing cluster geometry and retaining clinically meaningful multidimensional information. The full representation is therefore maintained to preserve the conceptual coverage of the proposed framework, while the ablation results provide evidence regarding which components most strongly influence the geometric properties of the clustering solution.

4.7. Cross-Algorithm Clustering Comparison

To examine whether the observed improvement associated with CWFF is consistent across different clustering algorithms, both feature representations—Approach B (equal-weight) and CWFF—were trained using the training data and evaluated exclusively on the held-out data within the $k = 2$ evaluation framework. For Agglomerative Clustering, which does not provide a built-in prediction mechanism for new observations, held-out cluster assignments were obtained using a nearest-centroid rule based on the centroids derived from the training data. The results for K-Means, GMM, and Agglomerative Clustering are presented in Table 10.

Table 10. Cross-Algorithm Comparison of Equal-Weight and CWFF Representations on the Held-Out Data ($k = 2$). Differences and 95% CIs in this table were obtained from a separate 500-iteration paired bootstrap applied to all three algorithms; therefore, the K-Means interval differs slightly from the 1,000-iteration estimate reported in Section 4.4 and the Abstract ([+0.014, +0.033]).

Algorithm	Silhouette (Equal-weight)	Silhouette (CWFF)	Difference (Bootstrap 95% CI)	Significant
K-Means	0.1980	0.2223	+0.024 [+0.015, +0.034]	Yes
GMM	0.1551	0.2586	+0.104 [+0.077, +0.131]	Yes
Agglomerative	0.1900	0.2142	+0.024 [+0.016, +0.031]	Yes

As shown in Table 10, CWFF produces higher Silhouette Scores than equal weighting across all three clustering algorithms. The largest difference is observed for GMM (+0.104; bootstrap 95% CI [+0.077, +0.131]), while both K-Means and Agglomerative Clustering show a difference of +0.024. The consistency of the direction of improvement across the three algorithms suggests that the observed gain is associated with the CWFF-weighted feature representation rather than being specific to a single clustering algorithm. The cross-algorithm comparison is also illustrated in Figure 4, which summarizes the held-out Silhouette Scores obtained with the equal-weight and CWFF representations.

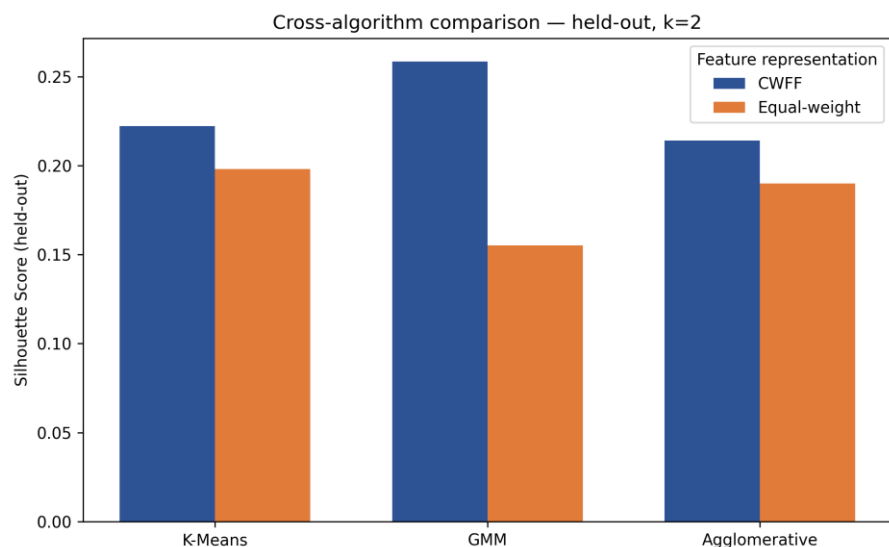


Figure 4. Comparison of held-out Silhouette Scores between CWFF and equal-weight representations across K-Means, GMM, and Agglomerative Clustering.

The analysis was further extended to HDBSCAN to examine a density-based clustering approach. Because HDBSCAN does not require a prespecified number of clusters, its default and training-selected configurations were evaluated separately. The results are presented in Table 11.

Table 11. HDBSCAN Performance Under Default and Training-Selected Settings on the Held-Out Data

Setting	Number of Clusters (Held-Out)	Training Noise (%)	Noise Held-Out (%)	Silhouette
Default (min_cluster_size = 70)	0	100.0	100.0	Cannot be calculated
Tuned on training data (min_cluster_size = 15)	3	30.0	33.7	0.238 (non-noise only)

As shown in Table 11, the two HDBSCAN configurations produce markedly different results. Under the default setting (min_cluster_size = 70), all observations in both the training and held-out data are classified as noise. Consequently, no clusters are formed and the

Silhouette Score cannot be calculated. This outcome reflects the clustering behavior under the specified parameter setting rather than an error in the analysis procedure.

When `min_cluster_size = 15` is selected using the training data, HDBSCAN identifies three clusters on the held-out data, with 33.7% of observations classified as noise. The Silhouette Score is 0.238 when calculated using non-noise observations only. The resulting three-cluster structure differs from the two-cluster solution obtained with K-Means, but this difference is not unexpected because the two methods use different principles for identifying cluster structure. K-Means assigns observations according to their proximity to centroids, whereas HDBSCAN identifies groups based on local data density.

Accordingly, the tuned HDBSCAN results provide a complementary view of the underlying data structure rather than a direct replication of the K-Means solution. The cross-algorithm results as a whole suggest that the effect of CWFF on geometric separation is maintained across several clustering paradigms, while the number and configuration of identified clusters can vary according to the underlying clustering assumptions.

5. Comparison

Previous studies on postpartum depression have predominantly used supervised learning approaches focused on binary classification or prediction. In contrast, the present study adopts an unsupervised clustering framework with semi-supervised feature weighting to identify heterogeneous risk subgroups. Because these approaches address different analytical objectives, their performance measures are not directly comparable on a common numerical scale. Accordingly, Table 12 provides a qualitative comparison based on the analytical paradigm, population and data characteristics, and research-domain coverage.

On the held-out data, CWFF achieved a Silhouette Score of 0.222, compared with 0.198 for equal weighting and 0.207 for PCA + K-Means. The difference between CWFF and equal weighting was statistically significant based on bootstrap evaluation. In addition, breastfeeding duration, which was used neither as a clustering feature nor to derive the CWFF weights, remained significantly different between the clusters on the held-out data ($p < 0.001$). The combination of held-out evaluation and independent validation provides evidence that the identified cluster structure is not evaluated solely on the same information used during model development.

Table 12. Qualitative Comparison of the Present Study with Related Studies

Study / Approach	Paradigm	Population & Data	Relative Difference from This Study
Ours (CWFF)	Unsupervised clustering with semi-supervised weighting	2,010 breastfeeding mothers from the UK; standardized questionnaire data	Held-out Silhouette Score of 0.222, with a significant improvement over equal weighting; independent validation using breastfeeding duration ($p < 0.001$); $k = 2$ determined using a pre-specified selection rule
Lin et al. [3] Fazraningtyas et al. [4], [21] ; Gul-mohamed et al. [18]; Garbazza et al. [14]	Supervised classification/prediction	Various cohorts; clinical and demographic indicators	Primarily provides binary risk prediction or a single risk score and does not explicitly characterize heterogeneous risk subgroups
Orchard et al. [12]	Unsupervised clustering of mental health profiles	General population, Ontario, Canada	Uses a comparable clustering perspective but is not specific to the breastfeeding domain and does not employ clinically guided feature weighting
Mathkor [24]	Integrated machine learning and statistical modeling	Maternal work stress, breastfeeding, mental health, and infant attachment	Provides closely related domain coverage but does not incorporate per-domain ablation analysis or the CWFF held-out evaluation scheme

The comparison in Table 12 places the present study within a complementary methodological perspective rather than directly competing with supervised postpartum depression

prediction studies. The primary distinction is the focus on subgroup characterization rather than binary risk prediction. The proposed framework further combines clinically informed feature weighting, prespecified cluster selection, held-out evaluation, and independent validation to examine whether the resulting subgroups remain interpretable beyond the information used to construct them.

6. Conclusions

Overall, this study supports the identification of two subgroups of breastfeeding mothers characterized by relatively higher and lower psychological burden using an unsupervised clustering framework with semi-supervised weighting and held-out evaluation. The selection of $k = 2$ was determined using a prespecified rule applied exclusively to the training data and subsequently evaluated on the held-out data. The resulting clusters showed consistent differences in psychological indicators and breastfeeding duration, providing a basis for descriptive characterization of the two subgroup profiles. Clinically Weighted Feature Fusion (CWFF) produced a significantly higher Silhouette Score than equal weighting on the held-out data, with scores of 0.222 and 0.198, respectively. This result indicates an improvement in geometric cluster separation under the proposed clinically informed representation. However, the difference in clinical separation based on η_{EPDS}^2 was not statistically significant. Thus, the current evidence supports a distinction between CWFF's contribution to geometric cluster quality and its additional contribution to separation based on clinical indicators. Independent validation using breastfeeding duration, which was used neither as a clustering feature nor to derive the CWFF weights, showed a significant difference between the clusters on the held-out data ($p < 0.001$). This finding provides additional evidence that the observed subgroup structure is not solely a consequence of the psychological variables incorporated into the CWFF weighting process. In contrast, maternal age did not differ significantly between the clusters, suggesting that the identified profiles are not similarly differentiated by this demographic characteristic.

Several limitations should be considered when interpreting these findings. First, the observed advantage of CWFF is primarily reflected in geometric cluster separation, while a corresponding improvement in clinical indicator separation has not been statistically established. Second, the analysis relies on a single UK dataset, and the generalizability of the identified subgroup structure to other populations and cultural contexts requires external validation. Third, the cluster labels are descriptive representations of empirical patterns and should not be interpreted as clinical diagnoses or established clinical categories. Future research should therefore evaluate the proposed framework in more diverse populations using external and longitudinal validation. Further investigation of alternative clinically informed weighting strategies may also help determine whether clinical separation can be strengthened while maintaining the stability and interpretability of the resulting cluster structure.

Author Contributions: Conceptualization: E.P.; Methodology: E.P. and A.F.; Formal analysis: E.P.; Investigation: E.P.; Data curation: E.P.; Validation: A.F. and P.; Writing original draft preparation: E.P.; Writing—review and editing: A.F. and P.; Visualization: E.P.; Supervision: A.F. and P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The dataset used in this study is publicly available through the Open Science Framework (OSF) repository at <https://osf.io/3hwrx/>. The dataset was originally collected by Braithwaite et al. and was reused for the analyses presented in this study. No new primary data were collected or generated. The analysis code (Python notebook) is available from the corresponding author upon reasonable request and will be deposited in a public repository upon acceptance.

Acknowledgments: The authors would like to thank Universitas Dian Nuswantoro and Universitas Darwan Ali for providing institutional support during the completion of this research. The authors also acknowledge the availability of the publicly accessible dataset through the Open Science Framework (OSF), which supported the analysis conducted in this study.

During the preparation of this manuscript, the authors used ChatGPT (OpenAI) for language editing, paraphrasing, and improving the clarity of the manuscript, and used ChatGPT to assist in designing and refining the visual layout of Figure 1 (workflow diagram). The authors reviewed and edited all AI-assisted content and take full responsibility for the accuracy, originality, and integrity of the final manuscript.

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

References

- [1] S. García-Méndez and F. de Arriba-Pérez, "Detecting and Explaining Postpartum Depression in Real-Time with Generative Artificial Intelligence," *Appl. Artif. Intell.*, vol. 39, no. 1, pp. 1–27, Dec. 2025, doi: 10.1080/08839514.2025.2515063.
- [2] E. M. Nagel *et al.*, "Maternal Psychological Distress and Lactation and Breastfeeding Outcomes: a Narrative Review," *Clin. Ther.*, vol. 44, no. 2, pp. 215–227, Feb. 2022, doi: 10.1016/j.clinthera.2021.11.007.
- [3] H. Lin *et al.*, "A Clinically Practical Postpartum Depression Predictor: Machine Learning Model Based on Simplified Indicators," *Research Square*, pp. 1–22, Aug. 01, 2025. doi: 10.21203/rs.3.rs-7079798/v1.
- [4] W. A. Fazraningtyas, B. Rahmatullah, H. Naparin, M. Basit, and N. A. Razak, "A predictive model for postpartum depression: ensemble learning strategies in machine learning," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 37, no. 1, p. 443, Jan. 2025, doi: 10.11591/ijeecs.v37.i1.pp443-451.
- [5] Z. Wang *et al.*, "Mapping global prevalence of depression among postpartum women," *Transl. Psychiatry*, vol. 11, no. 1, p. 543, Oct. 2021, doi: 10.1038/s41398-021-01663-6.
- [6] E. Hellyer *et al.*, "Postnatal Depression Beyond 12 Months: A Systematic Review and Meta-Analysis," *Int. J. Ment. Health Nurs.*, vol. 34, no. 2, pp. 1–13, Apr. 2025, doi: 10.1111/inm.70018.
- [7] H. Lei, Y. Yang, T. Zhu, X. Zhang, and J. Dang, "Network analysis of the relationship between non-suicidal self-injury, depression, and childhood trauma in adolescents," *BMC Psychol.*, vol. 12, no. 1, p. 234, Apr. 2024, doi: 10.1186/s40359-024-01729-2.
- [8] E. C. Braithwaite, A. Oftedal, A. Kaasen, Z. Ayorech, and M. Bekkhus, "A history of depression and prenatal depression are associated with a lower likelihood of breastfeeding initiation and maintenance, and more breastfeeding problems," *Arch. Womens. Ment. Health*, vol. 28, no. 1, pp. 139–146, Feb. 2025, doi: 10.1007/s00737-024-01479-5.
- [9] M. L. Gianni *et al.*, "Breastfeeding Difficulties and Risk for Early Breastfeeding Cessation," *Nutrients*, vol. 11, no. 10, p. 2266, Sep. 2019, doi: 10.3390/nu11102266.
- [10] R. Alimi, E. Azmoude, M. Moradi, and M. Zamani, "The Association of Breastfeeding with a Reduced Risk of Postpartum Depression: A Systematic Review and Meta-Analysis," *Breastfeed. Med.*, vol. 17, no. 4, pp. 290–296, Apr. 2022, doi: 10.1089/bfm.2021.0183.
- [11] S. I. Franklyn, J. Stewart, C. Beaupaire, E. Thaw, and R. J. McQuaid, "Developing symptom clusters: linking inflammatory biomarkers to depressive symptom profiles," *Transl. Psychiatry*, vol. 12, no. 1, p. 133, Mar. 2022, doi: 10.1038/s41398-022-01900-6.
- [12] C. Orchard, E. Lin, L. Rosella, and P. M. Smith, "Using unsupervised clustering approaches to identify common mental health profiles and associated mental health-care service-use patterns in Ontario, Canada," *Am. J. Epidemiol.*, vol. 193, no. 7, pp. 976–986, Jul. 2024, doi: 10.1093/aje/kwae030.
- [13] P. Vogazianos, E. Motrico, S. Domínguez-Salas, A. Christoforou, and E. Hadjigeorgiou, "Validation of the generalized anxiety disorder screener (GAD-7) in Cypriot pregnant and postpartum women," *BMC Pregnancy Childbirth*, vol. 22, no. 1, p. 841, Nov. 2022, doi: 10.1186/s12884-022-05127-7.
- [14] C. Garbazza *et al.*, "A machine learning model to predict the risk of perinatal depression: Psychosocial and sleep-related factors in the Life-ON study cohort," *Psychiatry Res.*, vol. 337, p. 115957, Jul. 2024, doi: 10.1016/j.psychres.2024.115957.
- [15] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, pp. 53–65, Nov. 1987, doi: 10.1016/0377-0427(87)90125-7.
- [16] M. Nedyalkova, S. Madurga, and V. Simeonov, "Combinatorial K-Means Clustering as a Machine Learning Tool Applied to Diabetes Mellitus Type 2," *Int. J. Environ. Res. Public Health*, vol. 18, no. 4, p. 1919, Feb. 2021, doi: 10.3390/ijerph18041919.
- [17] E. Prasetyaningrum, A. Z. Fanani, and D. W. Prabowo, "Identification of High-Risk Breastfeeding Mother Subgroups Through K-Means Clustering and Principal Component Analysis (Pca)," *Preprints*, Jun. 22, 2026. doi: 10.20944/preprints202606.1553.v1.
- [18] R. B. GulMohamed *et al.*, "AI Based Early Prediction of Postpartum Depression Using Multidimensional Maternal Health Data with Integrated Secure Data Storage and Privacy Protection Protocols," *J. Internet Serv. Inf. Secur.*, vol. 16, no. 1, pp. 548–569, Feb. 2026, doi: 10.58346/JISIS.2026.I1.031.
- [19] G. Rowles *et al.*, "Investigating the impact of breastfeeding difficulties on maternal mental health," *Sci. Rep.*, vol. 15, no. 1, p. 13572, Apr. 2025, doi: 10.1038/s41598-025-98357-6.
- [20] W. Zhang, K. Mao, and J. Chen, "A Multimodal Approach for Detection and Assessment of Depression Using Text, Audio and Video," *Phenomics*, vol. 4, no. 3, pp. 234–249, Jun. 2024, doi: 10.1007/s43657-023-00152-8.
- [21] W. A. Fazraningtyas, B. Rahmatullah, D. Dwi Salmarini, S. Arrieya Ariffin, and A. Ismail, "Recent advancements in postpartum depression prediction through machine learning approaches: a systematic review," *Bull. Electr. Eng. Informatics*, vol. 13, no. 4, pp. 2729–2737, Aug. 2024, doi: 10.11591/eei.v13i4.7185.
- [22] T. L. Galankin, A. Y. Bepalov, and H. Y. Moebius, "Clusterization of Behavioral and Psychological Symptoms of Dementia as Assessed by Neuropsychiatric Inventory: A Case Against the Use of Principal Component Analysis," *J. Alzheimer's Dis.*, vol. 98, no. 4, pp. 1483–1491, Apr. 2024, doi: 10.3233/JAD-231008.
- [23] D. Usman and I. Bin Mohamad, "Standardization and Its Effects on K-Means Clustering Algorithm," *Res. J. Appl. Sci. Eng. Technol.*, vol. 6, no. 17, pp. 3299–3303, Sep. 2013, doi: 10.19026/rjaset.6.3638.

- [24] D. M. Mathkor, “Explicating the interplay of maternal work stress, breastfeeding, mental health, and infant attachment using machine learning and statistical modeling,” *J. King Saud Univ. – Sci.*, vol. 38, no. 2, p. 15372025, Jan. 2026, doi: 10.25259/JKSUS_1537_2025.
- [25] E. Braithwaite *et al.*, “A mixed-methods dataset on maternal reported experiences of breastfeeding difficulties and mental health in the UK,” *Data Br.*, vol. 61, p. 111700, Aug. 2025, doi: 10.1016/j.dib.2025.111700.