Using Clustering Algorithms to Analyze Wearable Device Data to Support Pathways for Cultivating Medical Students' Humanistic Care Abilities

Jiankun Ren
School of Nursing, Hebi Polytechnic, Hebi 458030, Henan, China
Corresponding author: Jiankun Ren (e-mail: hlxyrjl1982@163.com)

ABSTRACT This study develops an unsupervised framework using Gaussian Mixture Models (GMM) to analyze multimodal physiological data from wearable devices, aiming to support the cultivation of medical students' humanistic care abilities. The framework integrates signals such as heart rate variability, skin conductance, and accelerometer readings obtained via textile-integrated sensors, and applies feature extraction and clustering to classify individual humanistic responsiveness. Personalized training pathways are generated based on the clustered physiological patterns, enabling targeted interventions ranging from foundational emotion awareness to advanced empathy simulation. Comparative evaluation demonstrates that GMM outperforms K-means, DBSCAN, hierarchical, and spectral clustering in terms of intra-cluster compactness, inter-cluster separation, and alignment with behavioral labels. Dynamic indicators, including low-to-high state transition frequency, provide better correlation with clinical performance than static measures, supporting flexible and adaptive training strategies. This approach offers a data-driven, end-to-end methodology for precise medical humanities education and can be integrated with antenna-enabled wearable devices, wireless physiological monitoring platforms, and edge-computing systems to facilitate real-time analysis and personalized educational interventions.

INDEX TERMS wearable devices, clustering algorithm, humanistic care, smart textiles

I. INTRODUCTION

HUMANISTIC care ability is one of the core competencies of medical education [1-2], which covers multiple dimensions such as empathy, communication, respect and altruism, and profoundly affects the quality of doctor-patient relationship and clinical care effect. However, its assessment has long relied on subjective questionnaires or observation scores In recent years, the development of wearable devices has provided a new approach to this problem. Particularly, the emergence of smart textiles allows for the seamless integration of sensors into medical gowns, enabling long-term, non-invasive monitoring without interfering with clinical operations. By continuously collecting physiological signals such as heart rate variability (HRV) and electrodermal activity (EDA), it is possible to objectively reflect an individual's emotional and empathetic responses in interpersonal interactions [3]. However, there is still a lack of reliable and interpretable computational models for how to effectively map these low-dimensional physiological data to complex and high-dimensional clinical humanistic behaviors.

Existing research mainly explores this mapping relationship from two levels. At the level of basic representation, scholars have verified the sensitivity of various physiological signals to humanistic states: for example, a tutorial on the review and analysis of HRV indicators [4]. EDA can be used to identify acute stress states and prevent acute stress from inducing various neurological dysfunctions [5]. PPG (Photoplethysmography) sensors can be used to accurately, instantly, and non-invasively measure vital signs, thereby reflecting overall health status and long-term well-being in real time [6-7]. In the textile field, researchers have developed various flexible strain and electrophysiological sensors that can be woven directly into fabrics, providing a stable interface for capturing these subtle humanistic-related physiological changes. A detailed application to multimodal emotion analysis is given to analyze physiological signals [8]. These studies collectively confirm that physiological signals can serve as potential proxy indicators of humanistic abilities, but a unified classification framework or a feasible educational intervention path has not yet been formed.

At the computational level, researchers have tried using supervised learning or simple clustering methods. Some scholars have used wearable devices to collect physiological signals and human motion data simultaneously to detect stress [9]. Others have proposed a label-free physiological state classification framework based on HRV, combined with unsupervised machine learning techniques to achieve a breakthrough in traditional detection systems [10]. Mandia used machine learning to predict students' participation ratings in simulated tasks [11]. Ran provided a comprehensive explanation of hierarchical clustering, selecting different methods for different types of data and scenarios [12]. Most existing methods rely on manually labeled behavioral tags, which are difficult to generalize, the models are relatively simple, and they lack a fine characterization of group heterogeneity. We position this study as focusing specifically on the physiological correlates of humanistic care expression. The derived 'high/low humanistic response' types are defined strictly as physio-behavioral phenotypes based on autonomic regulation patterns during structured interactions, serving as proxy indicators rather than a comprehensive assessment of the multifaceted socio-psychological construct. The core of the proposed framework is an unsupervised discovery of physiological patterns, independent of behavioral scoring during model training. Existing behavioral labels are used solely for post-hoc interpretation and validation of the clusters' external relevance, not for supervision.

This paper introduces an unsupervised Gaussian Mixture Model (GMM) framework designed to process multimodal data compatible with future intelligent medical apparel that objectively classifies medical students' humanistic care abilities and generates personalized training pathways. Leveraging publicly available wearable datasets, the framework integrates multidimensional physiological features to achieve fine-grained clustering of humanistic response patterns. Through behavioral semantic mapping and ruledriven mechanisms, it automatically produces differentiated intervention suggestions. The system completes an end-to-end transformation from raw physiological signals to actionable educational recommendations without new data collection, establishing a data-driven, reproducible paradigm for scalable medical humanities education.

II. EXPERIMENT AND METHODS SECTION

A. DATA SOURCES AND MODEL ARCHITECTURE

This study used three publicly available multimodal physiological datasets—WESAD, NSPWS, and YAAD (Young Adult's Affective Data - ECG and GSR Signals)—as data sources for wearable devices. These datasets all contain synchronous physiological signals and behavioral annotations collected in controlled or seminatural settings, suitable for simulating the physiological response characteristics of medical students in clinical communication or humanistic interaction scenarios. Table 1 summarizes the information of the collected publicly available datasets.

TABLE 1. Summary of Publicly Available Dataset Information

Dataset	Task Type	Sensors	Behavioral Labels	Samples Included
WESAD	Simulated doctor–patient interaction, stress-relief dialogue	PPG, EDA, 3-axis ACC (Accelerometer)	Baseline, Amusement, Stress	15
NSPWS	Continuous stress monitoring of nurses in real hospital settings	PPG, Temperature	Self-reported stress	28
YAAD	Emotion elicitation via affective videos	ECG, GSR	Self-annotated Valence, Arousal	27
Total	—	—	—	70

Datasets were screened for interpersonal interaction, emotion induction, or structured interview tasks: WESAD used Trier Social Stress Test (TSST) and relaxation phases; NSPWS extracted real-world clinical stress responses; YAAD employed emotion-eliciting video clips. These scenarios reliably induced humanistic care processes (emotion regulation, engagement, stress response). Subjects underwent rigorous exclusion for incomplete tasks, sensor faults, or missing labels. The final cohort comprised 70 ethically approved participants with informed consent, clustered via Gaussian mixture model into low (n=18), moderate (n=20), and high (n=32) humanistic response groups. Technical inclusion required photoplethysmography signals and behavioral labels for heart rate variability extraction and cluster validation. Multimodal data were timestampsynchronized to ensure temporal consistency. This subset provides ecologically valid, physiologically complete data with behavioral grounding for modeling objective humanistic care pathways.

These tasks (TSST, affective videos) were selected as standardized, ethically feasible proxies that elicit core psychophysiological processes relevant to clinical humanistic interactions—specifically, emotion regulation, empathic engagement, and stress response. While not identical to actual doctor-patient encounters, they provide a controlled foundation for modeling the physiological correlates of humanistic responsiveness. An unsupervised analytical framework based on Gaussian mixture models is constructed as shown in Figure 1:

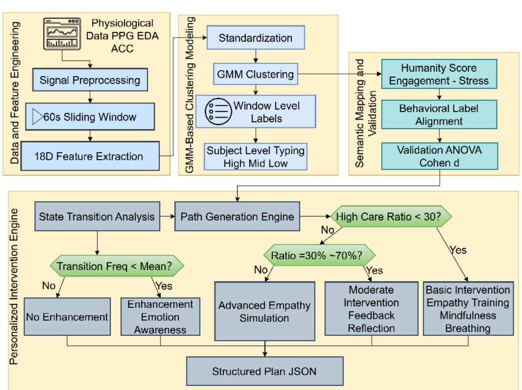

FIGURE 1. Overall Model Architecture Diagram

AEM JOURNAL.ORG

Figure 1 outlines the end-to-end framework comprising four sequential modules. Raw multimodal wearable signals PPG EDA and ACC undergo preprocessing window segmentation and extraction of 18 physiological features standardized to mitigate individual baselines. These features feed into Gaussian Mixture Model clustering generating window-level labels aggregated to subject-level humanistic response classifications high medium low. Semantic validation aligns cluster assignments with behavioral labels via Humanity Score Engagement minus Stress confirmed statistically through ANOVA. The personalized intervention engine then activates dual pathways: high-care state proportion triggers tiered training empathy modules mindfulness breathing communication feedback and advanced simulation while low-to-high state transition frequency below group average activates emotion awareness enhancement. All interventions integrate into structured JSON outputs translating physiological patterns into actionable educational strategies establishing a datadriven paradigm for medical humanities training.

The list of 18 physiological features extracted based on a 60-second non-overlapping sliding window is shown in Table 2 below. Covering heart rate variability, skin conductance, respiratory proxy indices, motion artifact control, and derived features [13-14], constituting the physiological feature space for unsupervised subtyping of humanistic care ability.

TABLE 2. List of 18 Physiological Characteristics

Category	Feature	Computation	Physiological Interpretation
HRV Time-domain	RMSSD	Root mean square of successive NN (Normal-to-Normal) interval differences	Parasympathetic activity
HRV Time-domain	pNN50	Percentage of successive NN intervals differing by > 50 ms	Heart rate variability (HRV)
HRV Time-domain	MeanNN	Mean of NN intervals	Baseline heart rate
HRV Frequency-domain	LF (Low Frequency) (0.04–0.15 Hz)	Power spectral density via Welch's FFT (Fast Fourier Transform)	Sympathetic–parasympathetic co-regulation
HRV Frequency-domain	HF (High Frequency) (0.15–0.4 Hz)	Power spectral density via Welch's FFT	Parasympathetic activity
HRV Frequency-domain	LF/HF ratio	LF power / HF power	Autonomic balance
EDA	SCL mean	Mean skin conductance level	Tonic arousal
EDA	SCR (Skin Conductance Response) peak count	Number of skin conductance responses per window	Phasic arousal events
EDA	SCR rise time (mean)	Mean time from onset to peak of SCR	Emotional response latency
Respiratory Proxy	Respiratory Rate	PPG-derived respiration frequency via amplitude modulation of PPG envelope	Breathing rate
Respiratory Proxy	RIIV (Respiratory-Induced Intensity Variation)	Standard deviation of PPG pulse amplitude within 60-s window	Respiratory variability & depth
Motion Artifact Control	ACC-RMS (Accelerometer Root Mean Square)	Root mean square of 3-axis accelerometer	Body movement intensity (covariate)
Derived Features	HR (Heart Rate) (bpm)	60 / MeanNN	Heart rate
Derived Features	SDNN (Standard Deviation of Normal-to-Normal RR Intervals)	Standard deviation of NN intervals	Overall HRV
Derived Features	VLF (Very Low Frequency) (0–0.04 Hz)	Very low-frequency power	Long-term regulatory processes
Derived Features	SCL variability	Standard deviation of SCL	Emotional stability
Derived Features	Respiratory Rate	PPG-derived respiration frequency	Breathing rate (Hz)
Derived Features	HRV Triangular Index	Histogram peak height / total count of NN intervals	HRV distribution morphology

These 18 features collectively model the psychophysiological processes underlying humanistic interactions: HRV indices reflect autonomic regulation crucial for stress modulation and engagement; EDA features index emotional arousal and reactivity; respiratory proxies are linked to emotion regulation; derived features offer integrated metrics of cardiovascular state. Crucially, ACC-RMS is included not as a direct indicator of care, but as a necessary covariate to control for motion artifacts that can confound physiological signals, thereby improving the validity of other care-related features.

HRV indices captured autonomic regulation: RMSSD and HF bands quantified parasympathetic activity, LF reflected mixed sympathetic-parasympathetic influence, and LF/HF ratio indicated autonomic balance. EDA features represented baseline arousal and event-related emotional responses. Respiratory proxies were derived from the ratio of dicrotic to main wave amplitude in the first derivative of PPG, leveraging cardiorespiratory coupling. ACC-RMS served as a motion artifact covariate. This multidimensional feature set comprehensively characterizes dynamic physiological states during doctor-patient interactions and stress regulation through integrated time-frequency domain analysis.

B. SIGNAL PREPROCESSING

This study adopted a uniform preprocessing procedure over the raw data to standardize both the temporal variation and features of multi-source physiological signals. All signals were resampled initially at 32 Hz to reach a compromise between computational efficiency and the physiological dynamic information. For the PPG signals, it employed fourth-order Butterworth bandpass filter [15-16]. This band preserves fundamental wave and its harmonic components of pulse wave and attenuates the baseline drift and high frequency motion artifacts [17-18]. The filtered signal was applied for pulse rate extraction and HRV analysis afterwards. For the EDA signal, it applied low-pass filter with cutoff value of 0.5Hz to attenuate EMG interference and high frequency noise while retaining EDA level and EDA component associated with sympathetic nerve activities. All the above filtering were implemented with zero phase [19-20], which would not cause phase distortion. Missing rates per signal channel were calculated. Segments with <5% missing data were filled via linear interpolation [21]; windows exceeding 5% missingness or exhibiting saturation/clipping were discarded. Only subjects retaining >90% valid windows were analyzed. For multimodal fusion, EDA and triaxial accelerometer signals were aligned to the PPG time axis using dataset-provided synchronization markers and linear interpolation, with alignment error constrained to ± 1 sampling point to ensure physiological consistency within temporal windows.

C. FEATURE ENGINEERING

In order to extract objective physiological indicators that are potentially related to the humanistic care ability of medical students from multimodal physiological signals collected by wearable devices, this study used a 60-second nonoverlapping sliding window to segment the preprocessed signal [22] and extracted a total of 18 features in each window. All feature calculations were implemented based on the NeuroKit2 toolkit and a custom Python script to ensure the reproducibility and physiological rationality of the method.

HRV features were derived from RR intervals detected in PPG signals. Time-domain metrics RMSSD pNN50 and MeanNN quantified parasympathetic activity. Frequency-domain indicators are calculated by Welch method fast Fourier transform to calculate power spectral density [23], extract the power of low-frequency and high-frequency bands, and calculate LF/HF ratio as a proxy indicator of autonomic balance. EDA features were decomposed into skin conductance level and response components using NeuroKit2 yielding mean SCL SCR peak count and mean SCR rise time to reflect baseline arousal emotional response frequency and response dynamics. Respiratory features were estimated via PPG-derived methods with RIIV defined as the standard deviation of PPG pulse amplitude within 60-second windows correlating with emotional regulation and parasympathetic activity. Triaxial accelerometer root mean square amplitude served as a motion artifact covariate to control for physical activity interference, and its calculation formula is as follows:

$$\mathrm{ACC}_{\mathrm{RMS}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(x_i^2 + y_i^2 + z_i^2\right)} \tag{1}$$

Including ACC-RMS as a covariate ensures that the extracted patterns of 'humanistic response' are attributable to autonomic and emotional regulation rather than extraneous physical activity.

All 18-dimensional features are z-score standardized within the window [24-25] and normalized at the subject level to eliminate individual baseline differences, ultimately forming a unified feature vector input clustering model.

D. CLUSTERING MODELING

To achieve an objective classification of medical students' humanistic care abilities, this study used Gaussian mixture model to perform unsupervised clustering on the extracted 18-dimensional physiological features [26-27]. The GMM assumes that the data is generated by a mixture of multiple Gaussian distributions, which can effectively characterize the continuity of physiological indicators and the heterogeneity between individuals. To determine the optimal number of clusters, we systematically evaluated K values from 2 to 6 using the Bayesian information criterion (BIC) to balance model complexity and fit, while simultaneously assessing intra-cluster compactness and inter-cluster separation through silhouette coefficients. Clustering was performed in a standardized feature space to eliminate dimensional effects on distance metrics. The Gaussian Mixture Model was trained using the Expectation-Maximization algorithm [28-29] with convergence tolerance of 1e-6 and maximum 100 iterations. To avoid local optima, k-means++ initialization was repeated 10 times, selecting the model with highest log-likelihood. Subject-level classification was determined by window-level cluster distributions: subjects were assigned to the cluster representing >50% of their windows, with boundary cases excluded. This approach successfully classified all 70 subjects into three distinct physiological response patterns, establishing a stable foundation for subsequent semantic mapping and personalized pathway generation.

To objectively analyze the effectiveness of Gaussian mixture model in humanistic care ability classification, this study uses silhouette coefficient, Bayesian information criterion, and algorithm comparison experiments to evaluate the system [30-31]. The silhouette coefficient for the i-th sample is defined as:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{2}$$

In equation 2, a(i) is the average distance between a sample i and all other samples in the same cluster, b(i) is the average distance between a sample iand all samples in the nearest dissimilar cluster, and the overall silhouette coefficient is the mean of all samples s(i), ranging from [-1, 1]. A larger value indicates a clearer clustering structure. The Bayesian Information Criterion (BIC) is defined for a GMM with K Gaussian components as follows:

$$\mathrm{BIC} = -2\log L + p\log N \tag{3}$$

$$p = K\left(d^2 + 3d + 1\right) - 1 \tag{4}$$

L is the maximum likelihood value of the model, N is the total number of samples, and p is the number of model parameters. The smaller the BIC value, the better the model achieves a balance between goodness of fit and complexity.

The algorithm comparison experiment, under the same feature input and preprocessing conditions, compared GMM, K-means, hierarchical clustering, spectral clustering, and DBSCAN (Density-Based Spatial Clustering of Applications with Noise) using silhouette coefficient, adjusted Rand index, and Calinski-Harabasz index.

Calinski-Harabasz index (CH index) [32-33]:

$$\mathrm{CH}(K) = \frac{\mathrm{Tr}(B_K)/(K-1)}{\mathrm{Tr}(W_K)/(N-K)} \tag{5}$$

In equation 5, Tr(BK) represents the trace of the inter-cluster scatter matrix, and Tr(WK) represents the trace of the intra-cluster scatter matrix. The larger the CH exponent, the higher the inter-cluster separation and the stronger the intra-cluster compactness.

Adjusting the RAND index to measure the consistency between clusters and formal groupings when reference labels are available [34-35]:

$$\mathrm{ARI}=\frac{\sum_{ij}\binom{n_{ij}}{2}-\left[\sum_i\binom{a_i}{2}\sum_j\binom{b_j}{2}\right]/\binom{N}{2}}{\frac{1}{2}\left[\sum_i\binom{a_i}{2}+\sum_j\binom{b_j}{2}\right]-\left[\sum_i\binom{a_i}{2}\sum_j\binom{b_j}{2}\right]/\binom{N}{2}} \tag{6}$$

In equation 6, nij represents the interaction size between cluster i and reference group j, and ai and bj represent their marginal totals. ARI ranges from [-1,1], with 1 indicating perfect agreement and 0 random matching. To evaluate alignment with ordinal behavioral labels, quadratic weighted Cohen's Kappa was additionally applied [28-29]. This weighting scheme assigns penalties proportional to the squared distance between misclassified categories, appropriately accounting for the ordered nature of humanistic response types where ordinary Kappa fails to distinguish severity of classification errors.Its formula is shown as follows.

$$\kappa=\frac{P_o-P_e}{1-P_e} \tag{7}$$

In equation 7, P_o represents the observed consistency ratio and P_e represents the expected random consistency ratio. When κ >0.6, it indicates that the clustering results are substantially consistent with the behavioral labels.

E. SEMANTIC MAPPING OF HUMANISTIC CARE TYPES

To align the unsupervised clusters with behavioral semantics, we constructed a composite index termed the Humanity Score, defined as Engagement - Stress. This operationalization is grounded in medical education frameworks which posit that effective humanistic care integrates observable engagement with the management of internal stress. The Engagement and Stress scores were derived directly from the annotated behavioral labels (e.g., 'Amusement', 'Stress') within each 60-second window in the source datasets. The mean HESD was then calculated for all windows assigned to each cluster to interpret the cluster's behavioral tendency.

Humanity Score was calculated according to the behavioral labels in each dataset by using the behavioral labels of each 60s, and then the Humanity Score was summarized based on the labels of the clusters to which the subjects belonged (take the mean value as the value of the cluster). In order to validate the statistical significance of the above semantic mapping, one-way ANOVA was conducted on three groups of Humanity Score [36-37].

F. PERSONALIZED TRAINING PATHWAY GENERATION

This study developed a physiological signal-based intelligent system to enhance nursing students' humanistic care abilities. Students are categorized by high-focus state duration: below 30% (low-response), 30-70% (moderate-response), and above 70% (high-response). The thresholds of 30% and 70% for high-care state duration were set based on the tertile distribution of the clustered cohort in this study, serving as an initial, data-driven heuristic to categorize low, moderate, and high responders. This operational choice provides a transparent and reproducible rule for pathway generation in the current framework. Low-response students receive foundational training. Moderate-response students undergo advanced modules. High-response students engage in complex simulations. The system dynamically monitors low-to-high state transition frequency; below-average performers automatically receive emotion awareness enhancement via physiological feedback and emotion labeling exercises. All interventions are output as standardized JSON structures containing student ID, care type, high-state percentage, triggered modules, and specific action plans, enabling seamless integration with medical education platforms.

III. RESULTS AND DISCUSSION

A. CLUSTERING PERFORMANCE EVALUATION

This study systematically evaluated the performance of Gaussian Mixture Model (GMM) on the human response genotyping task based on multimodal physiological data from 70 medical students. In a crosssectional comparison with K-means, hierarchical clustering, spectral clustering, and DBSCAN, GMM performed well in three metrics: silhouette coefficient, adjusted Land index (ARI), and Calinski–Harabasz index, as shown in Figure 2.

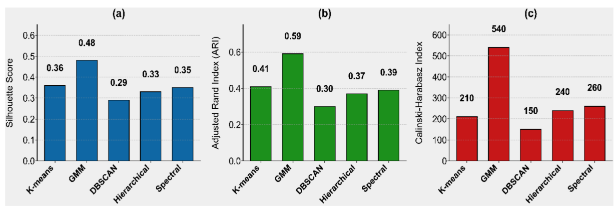

FIGURE 2. Performance Comparison of Different Clustering Models: (a) Silhouette Coefficient Comparison, (b) Adjusted Rand Index

Figure 2 compares the performance of GMM with four other clustering algorithms in classifying medical students' humanistic care abilities. GMM achieves the highest scores across all metrics: silhouette coefficient (0.48), adjusted Rand index (0.59), and Calinski–Harabasz index (540), outperforming K-means, DBSCAN, hierarchical clustering, and spectral clustering. These results demonstrate GMM's superior intra-cluster compactness, inter-cluster separation, and alignment with behavioral labels, confirming its effectiveness in capturing humanistic response patterns.

Gaussian Mixture Models excel in capturing humanistic care responses due to their alignment with physiological data characteristics: continuous transitions and individual heterogeneity produce non-spherical, overlapping distributions. GMM addresses this through soft clustering and explicit covariance structure modeling, avoiding rigid shape assumptions inherent in hard clustering methods like K-means. Its EM algorithm leverages inter-feature correlations to identify high-humanistic-response patterns. In contrast, DBSCAN proves sensitive to density variations in physiological windows, while hierarchical clustering fails to adapt to multimodal feature space structures.

In unsupervised clustering of complex, high-dimensional physiological data, a silhouette coefficient of 0.48, while moderate, indicates discernible structure and is substantially higher than comparative algorithms. More importantly, the primary validation of the GMM-derived clusters lies in their significant alignment with behavioral labels and their physiologically interpretable profiles, which collectively justify their use for generating hypothesis-driven educational pathways.

B. BEHAVIOR LABEL ALIGNMENT VALIDATION

To evaluate the consistency between different clustering algorithms and actual behavioral labels in the classification of medical students' humanistic care abilities, this paper constructs five confusion matrices and performs alignment analysis between the clustering results and the defined low, medium, and high behavioral groups. It is crucial to note that these behavioral labels were not used to guide the clustering process. They serve here as an independent, external benchmark to assess whether the data-driven physiological clusters correspond to meaningful behavioral differences, thereby evaluating the real-world relevance of the unsupervised discovery. The resulting confusion matrix comparison is shown in Figure 3.

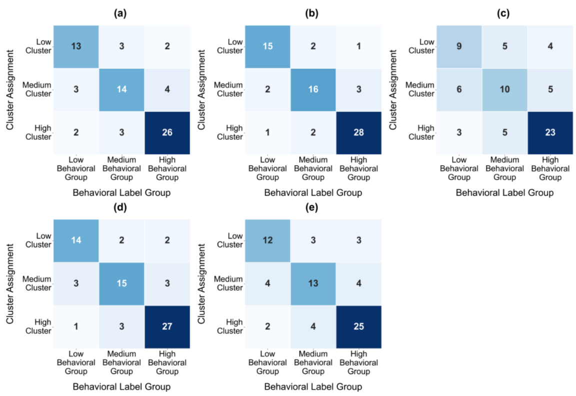

FIGURE 3. Comparison of Confusion Matrices for Each Model: (a) K-means Confusion Matrix; (b) GMM Confusion Matrix; (c) DBSCAN Confusion Matrix; (d) Hierarchical Clustering Confusion Matrix; (e) Spectral Clustering Confusion matrix

Figure 3 presents confusion matrices comparing clustering algorithms' alignment with reference behavioral labels. Rows correspond to algorithm-generated clusters while columns represent actual behavioral groups. Gaussian Mixture Model results in Figure b demonstrate strongest diagonal concentration with 15 16 and 28 correct assignments across groups. This model misclassifies only four high-behavior samples. In contrast DBSCAN results in Figure c show nine high-behavior misclassifications with diffuse off-diagonal patterns indicating poor semantic alignment. Gaussian Mixture Model's precise mapping confirms its effectiveness in capturing physiologically grounded humanistic response patterns.

Humanistic care ability manifests not as discrete categories but as continuous, individualized transitions. Gradual shifts in HRV and EDA during empathy escalation. Gaussian Mixture Models (GMMs) capture this complexity through probabilistic soft assignments and covariance modeling, aligning well with the gradientlike nature of behavioral labels. In contrast, hard clustering forces rigid boundaries, DBSCAN struggles with uneven physiological densities, and hierarchical clustering assumes tree-like structures ill-suited to multidimensional feature spaces. GMM thus offers both numerical superiority and conceptual coherence with the dynamic, probabilistic nature of humanistic competence, enabling reliable classification for downstream educational interventions.

C. ANALYSIS OF PHYSIOLOGICAL CHARACTERISTICS

To explore the physiological basis of different types of humanistic responses at the level of autonomic nervous system regulation, this study conducted inter-group differences in key physiological indicators among high, medium, and low humanistic response groups. All analyses were based on individual means after windowlevel feature aggregation. Independent samples t-tests were used to compare extreme groups, and Cohen's d effect size was calculated to assess the practical significance of the differences.

Figure 4 demonstrates physiological correlates of humanistic responsiveness identified by Gaussian mixture model clustering. Figure(a) reveals RMSSD increases with responsiveness level where high-responsiveness group median reaches 38 ms with minimal intergroup overlap. Figure(b) shows inverse skin conductance level trend with high group mean 1.01 significantly lower than low group mean 1.82 indicating superior emotional stability in high responders. Figure(c) validates this association via 1000-iteration label permutation test; observed Cohen's d effect size 1.20 exceeds null distribution p less than 0.001 confirming nonrandom physiological differentiation. Collectively these results establish objective autonomic signatures underlying humanistic care ability. These physiological phenotypes are interpreted as reflecting differential capacity for engagement and emotional regulation—key intermediate psychological processes underpinning humanistic care. Future work integrating explicit psychological measures would strengthen the construct validity of this mapping.

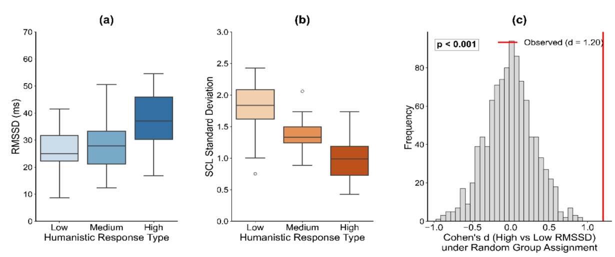

FIGURE 4. Box Plots of RMSSD and SCL Means for Three Types of Students: (a) Box Plot of RMSSD Means (b) Box Plot of SCL Means (c) Histogram of Zero Distribution for Randomization Test

This result reveals the physiological basis and quantifiable nature of humanistic care competence. High RMSSD reflects the parasympathetic nervous system's ability to exert control over the heart, which is closely related to empathic engagement, emotion regulation, and attentive listening. Low SCL variability indicates that the sympathetic system responds smoothly to interpersonal stress, making it less prone to emotional overload or disconnection. The reason why GMM can effectively identify this pattern is that its probabilistic modeling ability can capture the co-variant structure in the joint distribution of HRV and EDA, transforming fuzzy, continuous physiological dynamics into discrete subtypes with behavioral significance.

Association between State Transition and Clinical Manifestations Humanistic care capacity is not only reflected in static behavioral tendencies, but also relies on an individual's dynamic ability to regulate emotional and cognitive states during interactions. To test this hypothesis, this study constructs two types of indicators based on window-level GMM clustering labels:

(1) Static indicator: The percentage of time an individual spends in a "highly humanistic responsive" state;

(2) Dynamic indicators: the frequency of state transition from "low humanistic response type" to "high humanistic response type" during the task period (i.e. the number of effective adjustment events), which serves as a proxy for the flexibility of humanistic state;

The scatter plot shown in Figure 5 is obtained from the calculation:

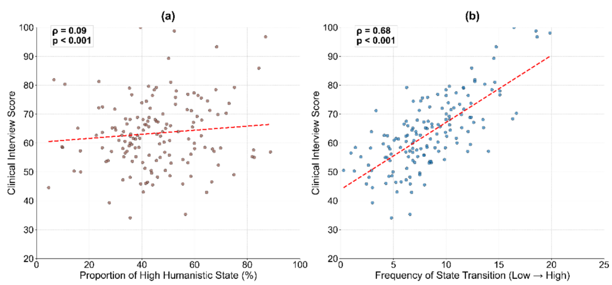

FIGURE 5. Scatter Plot of State Transition Frequency and Clinical Assessment: (a) Static Clinical Humanistic Performance Scatter Plot, (b) Dynamic Zero-bed Humanistic Performance Scatter Plot

Figure 5 reveals the difference in predictive value of static and dynamic physiological indicators for medical students' clinical humanistic performance through two scatter plots. (a) The figure shows the static high humanistic state ratio index, which is weakly positively correlated with clinical interview scores (20–100 points) (Spearman $\rho = 0.09$, $p < 0.001$): as the high state ratio increases from 0% to 100%, the score only slowly increases from about 60 points to close to 70 points. The data points are highly dispersed, indicating that the "high performance ratio" in a certain period is not enough to accurately reflect the overall humanistic ability. (b) The figure shows that the dynamic moderating indicator of "frequency of low to high empathic state transition" is significantly and strongly correlated with the score ($\rho = 0.68$, $p < 0.001$): the score jumps from about 45 points to 90 points, and the data points cluster more closely along the fitted line, indicating that frequent and flexible recovery from a low empathic state to a high empathic state better reflects the emotional regulation and humanistic response resilience of medical students in real interactions. This confirms that the essence of medical humanistic competence is not only about "whether one can show care," but also about "whether one can quickly return to a caring state after stress or distraction." Static proportion ignores the dynamic process and may misjudge occasional high performance as stable ability; while the frequency of state transition captures the flexibility of emotional regulation supported by the autonomic nervous system, which is the core of empathy exhaustion prevention and clinical communication resilience.

D. VERIFICATION OF THE RATIONALITY OF PERSONALIZED PATHS

To evaluate the educational rationality and discriminant validity of the generated personalized training pathways, the intervention plans automatically generated by the system for 70 participants were independently reviewed. The matching degree between the pathways and individual ability levels, the feasibility of the intervention measures, and the coherence of the educational logic were assessed. Further analysis of the distribution characteristics of the recommended content revealed that the system possesses a high level of category discrimination ability, resulting in the recommendation diagram shown in Figure 6.

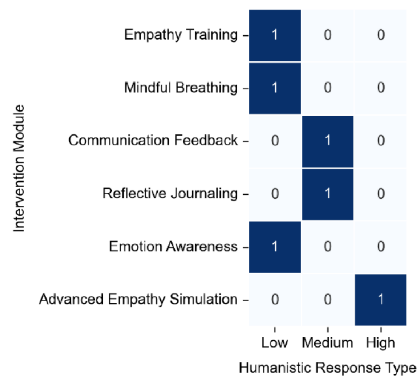

FIGURE 6. Student-Intervention Module Recommendation Chart

Figure 6 presents a binary recommendation matrix mapping three humanistic response types to six intervention modules. Low-response students receive foundational interventions: empathy training mindful breathing and emotional awareness to recognize and regulate emotions. Medium-response students receive communication feedback and reflection journaling for structured post-interaction reflection. High-response students receive only advanced empathy simulation to refine skills in complex clinical scenarios. This matrix directly translates physiological clustering results into operational modular educational pathways. Individuals with lower abilities need to first develop basic emotional awareness and empathy skills; those with intermediate abilities already possess the foundation, and the focus should be on transforming internal experiences into effective communication behaviors; those with higher abilities need to tackle more complex scenarios such as ethical dilemmas and cultural differences to avoid stagnation.

To investigate the differentiated recommendation strategies of the personalized training pathways generated in this study across different humanistic response types, Figure 7 shows the distribution of recommendation intensity across six intervention dimensions for medical students of low, medium, and high levels.

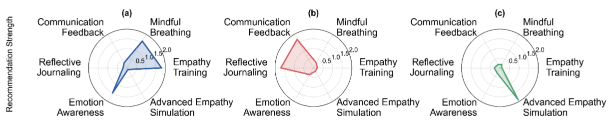

FIGURE 7. Radar Charts of Recommendation Strength for Three Types of Student Intervention Dimensions: (a) Radar Chart of Recommendation Strength for Low-response Intervention Dimensions, (b) Radar Chart of Recommendation Strength for Medium-response Intervention Dimensions, and (c) Radar Chart of Recommendation Strength for High-response Intervention Dimensions

Figures 7(a), 7(b), and 7(c) display graded intervention intensities across six modules for low, medium, and high humanistic response types. Low-response students show peak intensities in empathy training 1.8, mindful breathing 1.6, and emotion awareness 1.5, targeting foundational emotion regulation. Medium-response students exhibit strong focus on communication feedback and reflection journaling for post-interaction optimization while high-response students concentrate exclusively on advanced empathy simulation for complex clinical scenarios. Non-target modules maintain intensities below 0.4 across all groups confirming high pathway specificity. This visualization operationalizes physiological clustering into precise educational actions.

Low-ability students require foundational training in emotional awareness and empathy through mindfulness and emotional labeling. Intermediate-ability students focus on translating empathy into effective communication via feedback and narrative reflection. High-ability students advance through complex simulations involving cross-cultural and ethical dilemmas to build humanistic resilience. Figure 7 confirms this resource-efficient, tiered approach. The framework converts GMM physiological clusters into precise, interpretable educational pathways, transforming medical humanities education from experiential practice to a data-driven precision paradigm.

Table 3 . Compares of five clustering algorithms on the over all performance on the typing task of human response. Evaluation metrics include unsupervised profile coefficient, agree-ment with behavioral labels (Kappa, ARI), and classification Precision/Recall (macro-average). All models are applied on 18-dimensional normalized physiological fea -ture space.

TABLE 3. Summary Table of Quantitative Indicators

Metric	GMM (Proposed)	K-means	DBSCAN	Hierarchical clustering	Spectral Clustering
Silhouette Score	0.48	0.36	0.29	0.33	0.35
Cohen's Kappa (quadratic weighted)	0.62	0.48	0.35	0.41	0.43
Precision (macro-averaged)	0.75	0.61	0.52	0.58	0.6
Recall (macro-averaged)	0.75	0.59	0.5	0.56	0.58
Adjusted Rand Index	0.59	0.41	0.3	0.37	0.39

Table 3 shows that the proposed Gaussian Mixture Model—based on multiple distribution components— consistently outperforms other methods across all five evaluation metrics. It achieves the highest silhouette score, quadratic weighted Cohen's Kappa, adjusted Rand index, and macro-averaged precision and recall, indicating superior intra-cluster compactness, inter-cluster separation, and alignment with behavioral labels. By leveraging probabilistic soft assignments and modeling feature covariances, GMM effectively captures the continuous, multidimensional nature of humanistic care responses and individual heterogeneity. This yields more reliable and behaviorally meaningful groupings, enabling precise, individualized educational interventions.

This study employed standardized stress and emotion-induction tasks as proxies for clinical interactions. Although these tasks engage psychological and physiological systems central to humanistic care (e.g., arousal regulation, engagement), future research should validate the model using physiological data collected directly during real or simulated clinical encounters to strengthen ecological validity.

IV. CONCLUSION

This study establishes an unsupervised Gaussian Mixture Model (GMM) framework for classifying medical students' humanistic care abilities using multimodal wearable data. The study highlights the potential of using textile-based sensing interfaces to capture high-fidelity physiological signals in naturalistic clinical settings. GMM significantly outperformed K-means, DBSCAN, hierarchical, and spectral clustering across silhouette coefficient, adjusted Rand index, and Calinski-Harabasz metrics. Behavioral validation confirmed strong alignment between GMM-derived clusters and predefined groups, with only 18% misclassification in the highhumanistic group. Physiological analysis revealed high responders exhibited significantly elevated RMSSD and reduced skin conductance levels, indicating superior parasympathetic regulation and emotional stability. Crucially, dynamic state transition frequency showed strong correlation with clinical interview scores, surpassing static metrics and highlighting emotional flexibility as central to humanistic competence. The framework automatically generates personalized intervention pathways: low-responders receive empathy training, mindfulness breathing, and emotion awareness; medium-responders receive communication feedback and reflection journaling; high-responders undergo advanced empathy simulation. Expert reviews confirmed high rationality and group specificity of these pathways. This work delivers the first end-to-end data-driven paradigm for medical humanities education, transforming raw physiological signals into actionable, individualized training protocols. This research provides a theoretical and algorithmic foundation for the development of "Smart Medical Apparel," suggesting that the fusion of textile engineering and machine learning can significantly enhance the objectivity of medical humanities training.

AUTHOR CONTRIBUTIONS

The author is solely responsible for all aspects of this work, including study conception, design, execution, analysis, and manuscript preparation.

CONFLICTS OF INTEREST

The author affirm that he does not have any financial conflicts of interest.

FUNDING

This work was supported by Henan Medical Education Research Project (Project No.: WJLX2024127): Application Research of Virtual Simulation Platform in Rehabilitation Department Internship Teaching under the OBE Concept.

CONSENT TO PUBLISH

The manuscript has not been published before, and it is not being reviewed by any other journal. The author has approved the content of the paper.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author, upon request.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] D. Létourneau, J. Goudreau, and C. Cara, “Nursing students and nurses’ recommendations aiming at improving the development of the humanistic caring competency,” Canadian Journal of Nursing Research, vol. 54, no. 3, pp. 292–303, 2022, doi: 10.1177/08445621211048987.

[2] Л. Віннікова, “INTEGRATING HUMANISTIC VALUES INTO MEDICAL TRAINING: A CASE STUDY OF UK MEDICAL SCHOOLS,” Чернігівський колегіум" імені ТГ Шевченка, vol. 188, no. 32, pp. 138–143, 2025, doi: 10.58407/visnik.253223.

[3] G. Vos, K. Trinh, Z. Sarnyai, and M. R. Azghadi, “Generalizable machine learning for stress monitoring from wearable devices: A systematic literature review,” International Journal of Medical Informatics, vol. 173, no. 1, pp. 105026–105026, 2023, doi: 10.1016/j.ijmedinf.2023.105026.

[4] T. Pham, Z. J. Lau, S. A. Chen, and D. Makowski, “Heart rate variability in psychology: A review of HRV indices and an analysis tutorial,” Sensors, vol. 21, no. 12, pp. 3998–3998, 2021, doi: 10.3390/s21123998.

[5] A. Greco, G. Valenza, J. Lazaro, J. M. Garzon-Rey, J. Aguilo, C. de la Cámara, R. Bailon, et al., “Acute stress state classification based on electrodermal activity modeling,” IEEE Transactions on Affective Computing, vol. 14, no. 1, pp. 788–799, 2021, doi: 10.1109/TAFFC.2021.3055294.

[6] Z. Ebrahimi and B. Gosselin, “Ultralow-power photoplethysmography (PPG) sensors: A methodological review,” IEEE Sensors Journal, vol. 23, no. 15, pp. 16467–16480, 2023, doi: 10.1109/JSEN.2023.3284818.

[7] Neha, H. K. Sardana, R. Kanwade, and S. Tewary, “Arrhythmia detection and classification using ECG and PPG techniques: A review,” Physical and Engineering Sciences in Medicine, vol. 44, no. 4, pp. 1027–1048, 2021, doi: 10.1007/s13246-021-01072-5.

[8] A. Gandhi, K. Adhvaryu, S. Poria, E. Cambria, and A. Hussain, “Multimodal sentiment analysis: A systematic review of history, datasets, multimodal fusion methods, applications, challenges and future directions,” Information Fusion, vol. 91, no. 1, pp. 424–444, 2023, doi: 10.1016/j.inffus.2022.09.025.

[9] J. A. Castro-García, A. J. Molina-Cantero, I. M. Gómez-González, S. Lafuente-Arroyo, and M. Merino-Monge, “Towards human stress and activity recognition: A review and a first approach based on low-cost wearables,” Electronics, vol. 11, no. 1, pp. 155–155, 2022, doi: 10.3390/electronics11010155.

[10] G. Georgieva-Tsaneva, K. Cheshmedzhiev, Y. A. Tsanev, and M. Dechev, “Physiological State Recognition via HRV and Fractal Analysis Using AI and Unsupervised Clustering,” Information, vol. 16, no. 9, pp. 718–718, 2025, doi: 10.3390/info16090718.

[11] S. Mandia, R. Mitharwal, and K. Singh, “Automatic student engagement measurement using machine learning techniques: A literature study of data and methods,” Multimedia Tools and Applications, vol. 83, no. 16, pp. 49641–49672, 2024, doi: 10.1007/s11042-023-17534-9.

[12] X. Ran, Y. Xi, Y. Lu, X. Wang, and Z. Lu, “Comprehensive survey on hierarchical clustering algorithms and the recent developments,” Artificial Intelligence Review, vol. 56, no. 8, pp. 8219–8264, 2023, doi: 10.1007/s10462-022-10366-3.

[13] R. Ranjan, B. C. Sahana, and A. K. Bhandari, “Motion artifacts suppression from EEG signals using an adaptive signal denoising method,” IEEE Transactions on Instrumentation and Measurement, vol. 71, no. 1, pp. 1–10, 2022, doi: 10.1109/TIM.2022.3142037.

[14] D. Pollreisz and N. TaheriNejad, “Detection and removal of motion artifacts in PPG signals,” Mobile Networks and Applications, vol. 27, no. 2, pp. 728–738, 2022, doi: 10.1007/s11036-019-01323-6.

[15] A. Sujiwa and B. S. Maniani, “The Effect of Inaccurate Electronic Component Values on The Output Frequency Characteristics of Fourth Order Butterworth Type Low-pass Filter Circuits,” FARADAY, vol. 1, no. 1, pp. 1–8, 2025, doi: 10.33005/faraday.v1i1.4.

[16] A. A. Bisu, “Design and simulation of a 4th order high frequency bandpass filter for radar communication systems,” The Journals of the Nigerian Association of Mathematical Physics, vol. 66, no. 1, pp. 169–176, 2024, doi: 10.60787/jnamp-v66-324.

[17] F. Zhang, X. Tang, and L. Li, “Origins of baseline drift and distortion in Fourier transform spectra,” Molecules, vol. 27, no. 13, pp. 4287–4287, 2022, doi: 10.3390/molecules27134287.

[18] G. Müller and G. Sberveglieri, “Origin of baseline drift in metal oxide gas sensors: effects of bulk equilibration,” Chemosensors, vol. 10, no. 5, pp. 171–171, 2022, doi: 10.3390/chemosensors10050171.

[19] Y. Zuo, J. Mei, C. Jiang, X. Yuan, S. Xie, and C. H. Lee, “Linear active disturbance rejection controllers for PMSM speed regulation system considering the speed filter,” IEEE Transactions on Power Electronics, vol. 36, no. 12, pp. 14579–14592, 2021, doi: 10.1109/TPEL.2021.3098723.

[20] E. Grivel, B. Berthelot, G. Colin, P. Legrand, and V. Ibanez, “Benefits of zero-phase or linear phase filters to design multiscale entropy: Theory and application,” Entropy, vol. 26, no. 4, pp. 332–332, 2024, doi: 10.3390/e26040332.

[21] N. Zhang, K. Canini, S. Silva, and M. Gupta, “Fast linear interpolation,” ACM Journal on Emerging Technologies in Computing Systems (JETC), vol. 17, no. 2, pp. 1–15, 2021, doi: 10.1145/3423184.

[22] A. S. Nuran and S. Z. Sari, “GRADIENT BOOSTING APPROACH FOR MULTI-LABEL APPLIANCE STATE CLASSIFICATION IN NILM USING PUBLIC LOW-FREQUENCY ENERGY DATA,” Jurnal Media Elektrik, vol. 22, no. 3, pp. 221–230, 2025, doi: 10.59562/metrik.v22i3.9169.

[23] N. K. Mutlib, M. N. Ismael, and S. Baharom, “Damage detection in CFST column by simulation of ultrasonic waves using STFT-based spectrogram and welch power spectral density estimate,” Structural Durability & Health Monitoring, vol. 15, no. 3, pp. 227–227, 2021, doi: 10.32604/sdhm.2021.010725.

[24] C. Andrade, “Z scores, standard scores, and composite test scores explained,” Indian Journal of Psychological Medicine, vol. 43, no. 6, pp. 555–557, 2021, doi: 10.1177/02537176211046525.

[25] A. Jakkaraju, "Graph Neural Networks for Anomaly Detection in Cloud Infrastructure," Journal of Computer and Communications, vol. 13, no. 10, pp. 102–116, 2025, doi: 10.4236/jcc.2025.1310006.
[26] J. Tas, V. Rass, B. A. Ianosi, A. Heidbreder, M. Bergmann, and R. Helbok, "Unsupervised clustering in neurocritical care: a systematic review," Neurocritical Care, vol. 42, no. 3, pp. 1074–1086, 2025, doi: 10.1007/s12028-024-02140-w.
[27] M. Chaudhry, I. Shafi, M. Mahnoor, D. L. R. Vargas, E. B. Thompson, and I. Ashraf, "A systematic literature review on identifying patterns using unsupervised clustering algorithms: A data mining perspective," Symmetry, vol. 15, no. 9, pp. 1679–1679, 2023, doi: 10.3390/sym15091679.
[28] P. Jin, J. Huang, F. Liu, X. Wu, S. Ge, G. Song, D. A. Clifton, et al., "Expectation-maximization contrastive learning for compact video-and-language representations," Advances in Neural Information Processing Systems, vol. 35, no. 1, pp. 30291–30306, 2022. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2022/hash/c355566ce402de341c3320cf69a10750-Abstract-Conference.html.
[29] X. Zhu, K. Guo, S. Ren, B. Hu, M. Hu, and H. Fang, "Lightweight image super-resolution with expectation-maximization attention mechanism," IEEE Transactions on Circuits and Systems for Video Technology, vol. 32, no. 3, pp. 1273–1284, 2021, doi: 10.1109/TCSVT.2021.3078436.
[30] C. Sutherland, D. Hare, P. J. Johnson, D. W. Linden, R. A. Montgomery, and E. Droge, "Practical advice on variable selection and reporting using Akaike information criterion," Proceedings of the Royal Society B, vol. 290, no. 2007, pp. 20231261–20231261, 2023, doi: 10.1098/rspb.2023.1261.
[31] M. Agarwal, P. K. Tripathi, and S. Pareek, "Forecasting infant mortality rate of India using ARIMA model: A comparison of Bayesian and classical approaches," Stat Appl., vol. 19, no. 2, pp. 101–114, 2021. [Online]. Available: https://www.ssca.org.in/media/8_19_2_2021_SA_Sarla_Pareek_Final.pdf.
[32] W. Zhang, Z. Yue, J. Ye, H. Xu, Y. Wang, X. Zhang, and L. Xi, "Modulation format identification using the Calinski–Harabasz index," Applied Optics, vol. 61, no. 3, pp. 851–857, 2022, doi: 10.1364/AO.448043.
[33] Z. Syahputri and M. A. P. Siregar, "Determining the optimal number of k-means clusters using the calinski harabasz index and krzanowski and lai index methods for grouping flood prone areas in north sumatra," Sinkron: Jurnal dan Penelitian Teknik Informatika, vol. 8, no. 1, pp. 571–580, 2024, doi: 10.33395/sinkron.v9i1.13246.
[34] M. Sundqvist, J. Chiquet, and G. Rigaill, "Adjusting the adjusted rand index: a multinomial story," Computational Statistics, vol. 38, no. 1, pp. 327–347, 2023, doi: 10.1007/s00180-022-01230-7.
[35] A. D'Ambrosio, S. Amodio, C. Iorio, G. Pandolfo, and R. Siciliano, "Adjusted concordance index: an extension of the adjusted rand index to fuzzy partitions," Journal of Classification, vol. 38, no. 1, pp. 112–128, 2021, doi: 10.48550/arXiv.1509.00803.
[36] A. Ganesh, M. Ankesh, P. V. R. Reddy, G. Goyal, M. S. Thakur, and A. Jain, "One-way analysis of variance (ANOVA)," Vigyan Varta, vol. 5, no. 8, pp. 110–112, 2024, doi: 10.1007/978-3-030-64333-1_8.
[37] M. Alassaf and A. M. Qamar, "Improving sentiment analysis of Arabic Tweets by One-way ANOVA," Journal of King Saud University-Computer and Information Sciences, vol. 34, no. 6, pp. 2849–2859, 2022, doi: 10.1016/j.jksuci.2020.10.023.