RESEARCH ARTICLE

PCaLDI: Explainable Similarity and Distance Metrics Using Principal Component Analysis Loadings for Feature Importance

TAKAFUMI NAKANISHI, (Member, IEEE)

Department of Data Science, Musashino University, Koto-ku, Tokyo 135-8181, Japan

Comprehensive Research Organization, Waseda University, Shinjuku-ku, Tokyo 162-0041, Japan

e-mail: takafum.nakanishi@ds.musashino-u.ac.jp

ABSTRACT In the evolving landscape of interpretable machine learning (ML) and explainable artificial intelligence, transparent and comprehensible ML models are crucial for data-driven decision-making. Traditional approaches have limitations in distinguishing whether the observed importance of features in principal component analysis (PCA)-transformed similarity metrics is due to the intrinsic characteristics of the data or artifacts introduced by the PCA. This ambiguity hampers the accurate interpretation of feature contributions to similarity and distance metrics, which are fundamental to data-analysis techniques. To address these challenges, I introduce the novel PCA loading-dependent importance (PCaLDI), which elucidates the similarity and distance metrics by synergistically leveraging the strengths of PCA loadings and permutation feature importance. PCaLDI innovatively utilizes PCA loadings to prioritize the most influential features, streamlining the assessment of feature importance. This approach provides clearer insights into the contributions of the features and reduces the computational inefficiencies inherent to traditional methods. Importantly, PCaLDI uniquely clarifies the contributions of individual features to similarity metrics within the PCA-transformed space, distinguishing between the effects attributed to PCA and genuine influence of features on the similarity measures. This distinction is pivotal for accurately understanding the data structure and making informed decisions. Moreover, the versatility of PCaLDI extends to any data format compatible with PCA, highlighting its broad applicability and utility across data types. Comprehensive experiments and comparisons with baseline methods demonstrate that PCaLDI exhibits high effectiveness and efficiency, offering rapid and accurate assessments of feature importance with substantial reduced computational demands.

INDEX TERMS Explainable artificial intelligence (XAI), PCA loadings, PCA loading-dependent importance (PCaLDI), permutation feature importance, similarity/distance metrics.

I. INTRODUCTION

In the modern era, machine learning (ML) and artificial intelligence (AI) have become pivotal in various fields including medicine, economics, finance, and marketing [1], [2]. As these technologies increasingly influence decision-making, understanding the factors behind the predictions, estimates, and data features that shape these outcomes is essential. The roles of interpretable ML and explainable AI

The associate editor coordinating the review of this manuscript and approving it for publication was Lorenzo Ciani.

(XAI) are becoming increasingly important in identifying potential biases, confirming model reliability, and ensuring legal and ethical accountability.

Furthermore, media content is increasingly being analyzed in the vector form or when embedded representations are used. Assessing the similarity and distance of media content in these vector spaces can provide deeper insights into the meaning of the content. However, clarifying the reasoning behind the outcomes is crucial when evaluating the similarity or distance between two data instances. Specifically, introducing explainable similarity and distance metrics is vital
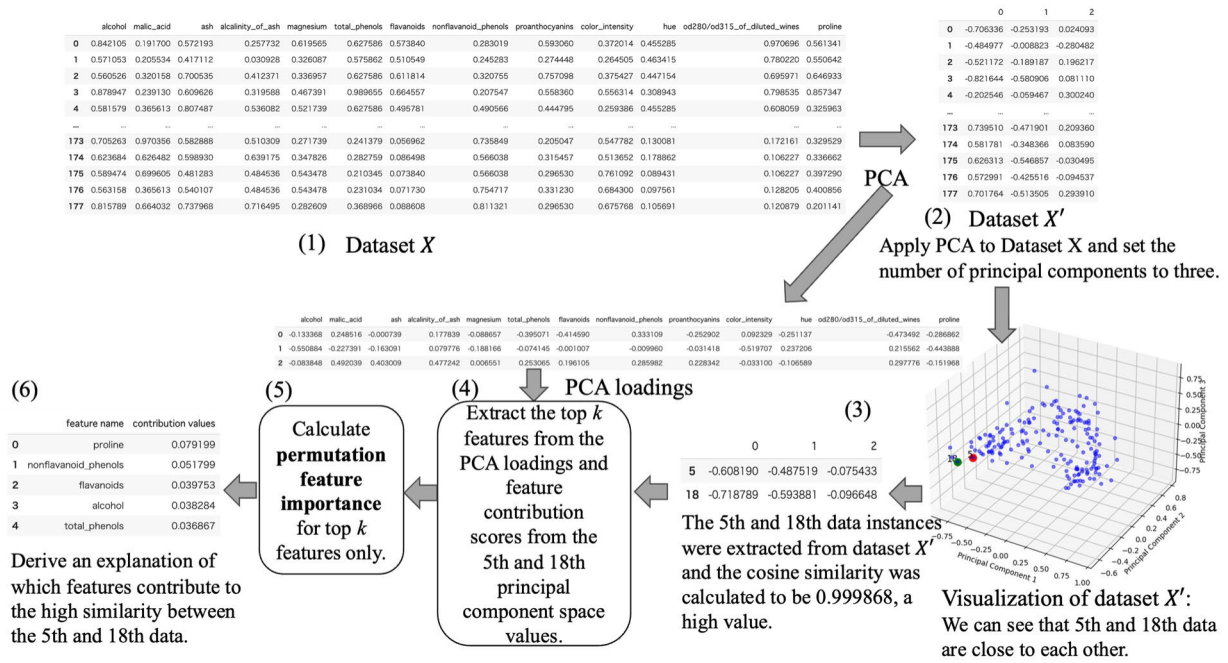
FIGURE 1. Overview of similarity computation using the proposed PCaLDI method on the Wine dataset [62]. Initially, (1) the Wine dataset is prepared as dataset *X*, (2) PCA is applied to the dataset *X*, (3) the data instances are selected from dataset *X*, and although it is possible to select data instances and measure the similarity and distance in this state, to reduce the curse of dimensionality, information redundancy, computational efficiency, and multicollinearity, we apply (2) PCA to generate dataset *X′* with dimensionality reduction. Next, (3) data instances (5th and 18th in the example) are obtained from dataset *X′*, and their similarity and distance are computed. In this example, the cosine similarity is 0.999868, which is a very high value. If we set the number of principal components to three, we can visualize dataset *X′* in a three-dimensional scatter plot and see that the 5th and 18th data instances are in proximity. To explore the reason for the high similarity between these two data instances, PFI is considered to evaluate the contribution of the features. However, if the number of features is large, the computation may not be completed within runtime. Therefore, I propose identifying important features in advance and calculating PFI only for these features. Specifically, we can calculate the feature contributions based on (4) PCA loadings and principal component space values for the 5th and 18th data instances. Subsequently, (5) PFI is computed for only the top k features in the feature contribution. Thus, we derive the features and their contributions to the similarity of the 5th and 18th data instances, as shown in (6). If the permutation feature importance calculation is applied to the data after PCA, the result represents the contribution of the principal component axis, which may lose its direct relevance to the original feature space. PCaLDI addresses this issue and provides a framework for evaluating the contribution to the original features, while still accounting for the PCA results, thereby improving both the interpretability and computational efficiency.

for interpreting algorithms such as *k*-nearest neighbors and *k*-means, which rely on these measures. Achieving explainable similarity and distance metrics is the first step toward a profound understanding of societal challenges.

This study introduces a novel method known as principal component analysis (PCA) loading-dependent importance (PCaLDI; Fig. 1). This method combines PCA loadings with permutation feature importance (PFI) to provide explainable similarity and distance metrics. PCaLDI effectively weighs feature contributions while minimizing the computational costs. Feature importance is evaluated using PCA loadings, and PFI is applied only to the top *k* features. Conventional PFI is computationally demanding, and the proposed approach significantly improves computational efficiency using PCA loadings as a preliminary step. PCaLDI is general, versatile, and applicable to any data format that supports PCA.

PCA loading is a well-known concept. However, its introduction into the description of similarity metrics may have the following effects: In conventional PFI, the values of the principal components in the principal component space are permuted, or permutation is performed in the feature space. PCA is performed, the similarity is weighed, and the

differences are calculated. In the former case, features that contribute to the original feature space cannot be obtained, and only the contributions of the principal components can be calculated. In the latter case, permutation, PCA, and similarity are performed on each feature; therefore, determining whether a difference exists or purely initial similarity is difficult using PCA. Therefore, in this method, by determining the *k* features that contribute to each principal component in the PCA loading, important features can be filtered out. This enables permutation to significantly reduce computational requirements without sacrificing the depth of interpretation provided by the importance technique. In addition, by deriving the features that contribute to the similarity via differences in the similarity permuted only to *k* features, the computational cost is reduced and features that contribute to the similarity, including PCA, can be reliably extracted.

The primary contributions of this study are as follows:

- It introduces PCaLDI, an explainable similarity/distance metric that combines PCA loadings with PFI. Traditional PFI methods are computationally demanding, especially on large datasets, because they require the importance of each feature to be computed across many

permutations. PCaLDI overcomes this important technical difficulty by reducing the computational complexity while maintaining interpretability.

- To validate the efficacy of PCaLDI, I implemented the method along with several baselines and conducted comparative experiments to demonstrate its effectiveness and utility.

The remainder of this paper is organized as follows: Section II reviews relevant previous studies. Section III elaborates on the formulation of PCaLDI. Section IV discusses the implementation of PCaLDI across various data formats, supported by comparative experiments. Section V presents a discussion on this study. Finally, Section VI concludes the paper.

II. RELATED WORKS

In recent years, substantial research has been conducted on ML interpretability and XAI, with numerous surveys and comprehensive descriptions [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22] addressing these issues. Chamola et al. [10] introduced the concept of trustworthy AI, providing a comprehensive review on building trustworthy AI and XAI systems. They emphasized the importance of transparent posterior models in developing XAI and outlined their potential drawbacks and pitfalls. The prominent XAI techniques in ML include LIME [23], SHAP [24], LRP [25], and Grad-CAM [26].

Various methodologies have been proposed to enhance explainability. These include partial-dependence plots [27], [28], which focus on explanatory variables and visualize the estimates and impact of each feature; individual condition explanations [29], which assess feature importance by randomly reordering or removing certain features; and leave-one-feature-out importance [30], [31]. Sensitivity analysis-based feature importance extraction methods [32] evaluate the effects of changing the input features across all subsets, considering the interactions and redundancies among the features. However, these techniques are often computationally intensive and expensive for high-dimensional datasets. For example, they require an evaluation of the impact of features on all subsets for n features. The quantitative input influence measure [33] is another approach aimed at quantitatively evaluating the impact of individual input variables on algorithmic predictions and estimations.

Aslam et al. [34] acknowledged that XAI recognizes the importance of human factors and the cognitive effort required to understand explanations. They proposed incorporating human factors into the design process and intentionally designing explanations that are aligned with the mental model of the user. In addition, they suggested the development of an ontology-driven conceptual model for XAI. Zhang et al. [35] emphasized the significance of XAI methods in cybersecurity applications. They presented a comprehensive survey of the up-to-date XAI approaches that are applicable to cybersecurity problems. Theissler et al. [36] focused on time-series data and reviewed various techniques for organizing and analyzing such data. They categorized these methods into point-, subsequence-, and instance-based approaches. They also identified open research areas for types of explanations and for assessing explanations and interpretability. Bobek et al. [37] focused on the XAI problem in unsupervised ML. They observed that global explanations might be overly broad, whereas local explanations based on the centers of gravity may overlook valuable information regarding the shape and distribution of clusters. They proposed a new approach known as "cluster analysis with multidimensional prototypes," which assists cluster-analysis experts by providing human-readable rule-based explanations.

I have previously developed a method known as approximate inverse model explanations (AIME) to visualize the difficulty of identifying features in black-box models. AIME constructs an approximate inverse operator for the ML model [38]. My experiments demonstrated that AIME provides simpler and clearer explanations than those generated by established methods, such as LIME [23] and SHAP [24]. The rationale behind such performance is the unique approach to interpreting black-box models. By addressing the inverse problem, AIME aligns more closely with human cognitive processes that seek to understand outcomes based on their causes. This method derives approximate inverse operators and establishes a direct link between input features and model predictions, thereby offering a more intuitive understanding of the decision-making process and rationale behind it. In addition, AIME introduces a representative instance similarity distribution plot, thereby enabling a unique visualization of the connection between the dataset and its predictions, which enhances interpretability. Although LIME and SHAP provide valuable local feature contributions to model predictions, the combination of global and local feature importance assessments of AIME facilitates a comprehensive understanding of the decision-making process of the model, which simplifies the explanations.

My concern extends beyond ML models to include the explainability of similarity and distance metrics between vector representations such as embedded representations. This is relevant for supervised ML as well as unsupervised methods such as k-means [39], in which understanding the calculations behind the similarity and distance metrics is crucial for explaining cluster memberships.

Several studies have investigated the extraction of features that contribute to similarity and distance metrics between datasets. For example, Ngu et al. [40] focused on image databases and developed a combination of multiple visual features that emulates human identification and classification of images using multiple visual elements. This approach essentially serves as a feature extraction method based on human visual features. Li [41] introduced a similarity measurement technique for symbol sequences based on order statistics by identity by state (IBS) and combined it with CCNN, which leverages convolutional neural network (CNN) feature extraction and IBS distance measurement to analyze movements in Sanda (or Sanshou, a form of Chinese martial

arts). Specifically, the CCNN model denoises the acquired images of movements and simplifies them into sequence diagrams representing human skeletal nodes, thereby offering a fresh perspective on feature selection for similarity metrics. Entezami and Shariatmadar [42] applied an autoregressive moving average model to select eigenmode functions for extracting model residuals as damage susceptibility features. They introduced a multivariate distance correlation method to identify structural damage by comparing multivariate residual sets from the normal and damaged states. Opitz et al. [43] proposed a method that uses abstract meaning-representation (AMR) graphs to introduce explainable metrics for assessing controversial similarities. This method uses AMR graph metrics to gauge the similarity of arguments and provides explanatory information for assessment. Fang et al. [44] employed patient similarity analysis to develop an explanatory and interpretable clinical risk prediction model. Specifically, they used the k-nearest neighbor method to model patient similarity based on expert input. Papoutsakis and Argyros [45] presented an unsupervised approach for estimating the similarity between two videos using the bipartite graph edit distance of the corresponding action graphs. This method accounts for the similarity between videos when the objects in the videos are semantically related, interact similarly, or both, by establishing spatiotemporal correspondences between objects across videos. Lin et al. [46] introduced explainable cosine (xCos), a new explainable similarity metric for face detection. xCos calculates the weighted sum of similarities by drawing upon the facial grid features extracted using a CNN feature extractor. These weights are determined from an attention map, thereby identifying facial features that are more representative.

PFI has been leveraged [47] to identify features that significantly influence similarity and distance calculations. However, the computational burden of PFI is considerable because of the repeated reconstruction of models while shuffling each feature.

König et al. [48] introduced the relative feature importance method as a more detailed alternative for assessing feature importance. This method encompasses PFI and conditional feature importance, and it is designed to evaluate the feature importance across any subset of features. This method not only clarifies the inter-feature relationships but also appears to reduce computational complexity. Meanwhile, Doyen et al. [49] proposed "hollow-tree super" (HOTS) as a novel approach for dissecting and visualizing feature importance in boosted tree models. They asserted that HOTS is particularly useful for managing large feature datasets. Niu et al. [50] introduced the batch-wise PFI (BPFI) method as a lightweight alternative to the traditional PFI. Instead of applying permutations across an entire test dataset, BPFI restricts the permutations to individual batches. Consequently, a computationally efficient approximation of the traditional feature importance metrics is obtained because the permutations are confined to smaller data subsets.

Previous research has focused on the extraction of important features, particularly by leveraging PCA and its loadings [51]. PCA transforms a large set of correlated features into a smaller set of uncorrelated variables known as principal components, which effectively captures the variability of the data. For instance, Baisantry et al. [52] introduced a feature selection method based on the loadings acquired from spectral-spatial principal components. Similarly, Li and Qiu [53] proposed an improved PCA-based feature selection method that uses the loadings of each indicator on various principal components, along with its variance contribution. Another study conducted by Baisantry et al. [54] suggested a novel band selection method based on PCA loadings. Their method aims to integrate the benefits of two-dimensional reduction strategies by selecting bands in relation to their associated principal components. This approach ensures the retention of the original bands, thereby challenging the prevailing notion that all essential information is consolidated in the upper principal components. My findings indicate that the middle principal components play a more dominant role in discrimination, particularly when different classes are spectrally similar or overlapping. Kiyoki et al. [55] used eigenvectors derived from eigenvalue decomposition to create a vector space for a semantic associative search. They introduced the semantic spectrum analyzer, which enhances search results by considering the variability in the individual values of the eigenvectors. Similar to this method, my approach incorporates PCA loadings and emphasizes the individual values of eigenvectors.

The method proposed in the present study innovatively combines PCA loadings with PFI to offer a fresh perspective on the similarity and distance metrics. The calculation of the PFI is typically computationally complex. However, by initially employing PCA loadings to assess feature importance, the permutation calculations can be limited to only the top-ranking features. This considerably reduces computational costs, especially when dealing with large datasets or high-dimensional feature spaces.

I further validated the efficacy of this approach by comparing it with several established baseline methods, the details of which are presented in Section IV. A key strength of this method is its versatility; it can be applied to different types of data, such as tabular or textual data, to extract features that influence the similarity or distance between data instances.

III. FORMULATION OF PCaLDI

This section outlines the framework of the proposed PCaLDI. Section III-A presents the key principles and theoretical underpinnings of PCaLDI as well as its limitations. Section III-B elaborates on the specific algorithms developed for PCaLDI, based on these conceptual foundations. Section III-C assesses the computational complexity of the PCaLDI and its suitability for large datasets.

A. INTRODUCTION TO PCaLDI

The goal of working with a dataset is to quantify the similarity or distance between two instances. In the context of similarity, a high value indicates that the instances are similar. Conversely, a small distance indicates proximity. However, the straightforward calculation of similarity and distance in a dataset poses several challenges. In particular, the "curse of dimensionality" (a term first coined by Bellman [56]) describes the phenomenon in which the distinction among distances between data points becomes less clear as the data dimensionality increases.

One strategy for solving such problems is to introduce PCA as a preprocessing method. This method mitigates the dimensionality problem and reduces noise by condensing the main data variations into lower dimensions. The compact, low-dimensional representation also improves computational efficiency and enables the visual confirmation of data relationships when two or three principal components are used.

However, limitations exist in contexts in which PCA is less effective, for example, nonlinear data structures, features with uniform variance, noisy data, sparse data, and data that contain many outliers. In such situations, alternative dimensionality reduction methods such as t-SNE [57] or UMAP [58] may be more suitable. Nevertheless, this study takes advantage of the unique properties of PCA loadings for PCA-enabled datasets.

A key strength of PCA is its computational loading, which clarifies the relationship between each variable and its corresponding principal components in the original feature space. These PCA loadings serve as indicators of feature importance, thereby facilitating the identification of important features for determining similarity. Distance PCA loading represents the contribution of a variable (feature) to each principal component extracted via PCA in the matrix form. This matrix indicates the extent to which the variance in the dataset is explained by a particular variable and is therefore key to understanding how each feature shapes the overall data structure. For similarity measures in a space consisting of principal component axes, features with high PCA loading values play an important role in shaping similarity within the dataset. Specifically, through the weighting indicated by the PCA loading, the importance of features in a particular principal component can be quantitatively evaluated, thereby efficiently identifying the features that contribute the most to similarity.

Therefore, I introduce PCaLDI, a novel metric framework for assessing similarity and distance; this framework combines the strengths of PCA loadings and PFI. This synergistic combination enables the evaluation of both the global variability within a dataset and feature-specific effects on variables of interest, such as similarity and distance. By performing permutations on each of the k features, features that are particularly important in the case of this similarity measure can be derived, and the computational cost can be significantly reduced compared with that of permuting all

features. As opposed to conventional PFI, the contribution of features in the original feature space to the similarity in the PCA space can be obtained via PCA loading. In conventional PFI, the difference is obtained either by permuting the values of the principal components in the principal component space or by permuting in the feature space, performing PCA, weighing the similarity, and then obtaining the difference. In the former case, the contributing features in the original feature space cannot be obtained, and only the contribution of the principal components can be calculated. In the latter case, permutation, PCA, and similarity are performed on each feature, which makes it difficult to distinguish between a difference in the PCA or purely initial similarity. Therefore, in this method, by determining k features that contribute to each principal component in the PCA loading, important features can be filtered. In addition, by deriving the features that contribute to similarity via differences in the similarity permuted only to k features, the computational cost is reduced and features that contribute to the similarity, including PCA, can be extracted more reliably. In other words, PCaLDI, which combines PCA loadings and PFI, provides new insights into the similarity measurement in the PCA space with a lower computational cost for deriving which features initially contribute to the similarity. This approach is a mathematically and statistically justified means of clarifying the structure of the data and the features that are important in shaping the similarity, while significantly reducing the computational cost compared with that of traditional PFI. It is applicable to different data types and provides transparent and interpretable similarity and distance metrics. In this section, I will describe the mathematical formulations for PCA, PCA loadings, and PFI.

1) PCA

Let X denote the original dataset that is represented as a matrix with m instances (rows) and n features (columns). The mean and standard deviation are obtained for each feature in the original dataset X. Subsequently, normalization is performed by subtracting the mean and dividing it by the standard deviation for each feature, as expressed in the following equation:

$$x_{i,j} = \frac{x_{i,j} - \mu_j}{\sigma_j}, \qquad (1)$$

where $x_{i,j}$ is the element in row i and column j of dataset X, μ_j is the mean of column j, and σ_j is the standard deviatio of column j. Next, we calculate the covariance matrix, C, using the normalized dataset X:

$$C = \frac{1}{m} X^T X, \qquad (2)$$

where T denotes the transpose operation, and m is the total number of data points in the dataset, that is, the total number of rows of data. Subsequently, we proceed to determine the eigenvalues and corresponding eigenvectors of C.

$$C v = \lambda v, \qquad (3)$$

where λ represents the eigenvalue, and v is the corresponding eigenvector. A maximum of n eigenvalues λ and eigenvectors v (number of features) can be derived. The eigenvalues and eigenvectors correspond to one another, and the eigenvalues represent the variance of the corresponding eigenvector, which serve as the principal components i PCA. In particular, the eigenvector corresponding to the largest eigenvalue is the first principal component. These eigenvectors are arranged in the order of increasing eigenvalues. Subsequently, when the number of principal components p is set to two, if the eigenvectors are selected in the order of the largest eigenvalues, they become the first and second principal components. Finally, dataset X is projected onto the first and second principal component. In essence, the dataset can be reduced to p dimensions. Specifically, if X' represents the dataset after dimensionality reductio to p dimensions, the following equation applies:

$$X' = XV_p, \tag{4}$$

where V_p is a matrix consisting of the first p principal components as the column.

2) PCA LOADINGS

PCA loadings quantify the relationship between each original feature and the associated principal components. They serve as the markers of the extent to which each variable contributes to a given principal component. Each element of the eigenvector is considered a loading for the principal component of that variable. Mathematically, the loading for each feature on the ιth principal component derived from the covariance matrix of dataset X can be expressed as

$$l_\iota = \sqrt{\lambda_\iota} v_\iota, \tag{5}$$

where λ_ι and v_ι represent the ιth eigenvalue and eigenvector, respectively, and v_ι is the ιth principal component. l_ι ranges from -1.0 to 1.0. This equation (5) scales each element of the eigenvector using the square root of its corresponding eigenvalue. Each element of vector l_ι represents the loading of each feature on the principal component. These loadings offer insights into the relative importance or influence of each variable by showing its contribution to a specific principal component [59].

3) PFI FOR SIMILARITY/DISTANCE

PFI is a technique for assessing the influence of a given feature on the predictive power of an ML model. The feature values are randomly shuffled within the dataset, and the change in prediction quality is monitored. A significant variation implies that the feature is crucial for the model predictions. Although PFI is commonly used to evaluate model performance, in this study, the model is adapted to assess the contribution of each feature to the calculation of the similarity or distance metric. This study introduces a novel framework for directly assessing the importance of features in calculating similarity and distance, thereby distinguishing

it from traditional model-based predictions. This is achieved using the following formulation:

PCA is performed on the original dataset X to obtain the transformed dataset X', as detailed in Section III-B. The similarity or distance ρ between two instances x'_τ and x'_υ of dataset X' is calculated as

$$\Delta_{original} = \rho\left(x'_\tau, x'_\upsilon\right). \tag{6}$$

A new dataset $X_{permuted}$ is generated by shuffling (replacing) the values of feature j in X. Subsequently, PCA is reapplied to $X_{permuted}$ to produce $X'_{permuted}$. The data instances $x'_{\tau\ permuted}$ and $x'_{\upsilon\ permuted}$ that are replaced and transformed by this operation are used to calculate the new similarity or distance as

$$\Delta_{permuted} = \rho(x'_{\tau\ permuted}, x'_{\upsilon\ permuted}). \tag{7}$$

The importance of feature j, denoted by I_j, is measured as the difference between the original and new similarity or distance metrics as

$$I_j = \left| \Delta_{permuted} - \Delta_{origial} \right|. \tag{8}$$

A large value signifies that feature j has a substantial impact on the calculation of similarity or distance. This procedure is applied iteratively to all features to evaluate their individual importance.

B. PCaLDI ALGORITHM

As detailed in Section III-B, the computation of PFI for similarity or distance is computationally demanding. This is particularly true for datasets with numerous features, where rea-time derivation is infeasibl.

PCaLDI can address this limitation. Using this approach, we focus only on the top k features ranked by their contribution levels. The steps involved are outlined below.

The feature contribution is computed by integrating the loadings obtained via PCA with the projected values of the two data instances under consideration. Specifically, let x'_τ and x'_υ represent two data instances projected onto the principal component space. The corresponding contribution scores, denoted by s, are calculated as follows:

$$s = \left| x'_\tau \cdot L \right| \circ \left| x'_\upsilon \cdot L \right|, \tag{9}$$

where s is an n-dimensional vector that encapsulates the contribution scores of features in the original dataset. Both x'_τ and x'_υ are p-dimensional vectors representing the data instances following the PCA transformation for rows τ and υ, respectively. L is an $n \times p$ matrix of the loadings, where each column corresponds to the loading vector for a specific principal component, $\circ$ represents the Hadamard product (element-wise multiplication), and $\cdot$ denotes the inner product in the p-dimensional real space. Using this formulation, the contribution scores for each feature can be effectively computed. This approach essentially calculates a score to gauge the influence of each feature on the similarity or distance between x'_τ and x'_υ. This can be achieved by combining the

post-PCA data values with the loadings to form the contribution vector s.

Thereafter, the contribution score vectors s are sorted by magnitude, and the PFI is calculated for only the top k features based on a predetermined k. Subsequently, we determine the extent to which each feature contributes to the calculated similarity or distance.

This approach enables us to derive the contributions of the features from multiple perspectives, thereby minimizing the computational cost of the PFI. Additionally, this methodology underscores the adaptability and broad applicability of PCaLDI to various PCA-enabled data formats.

The applicability and effectiveness of the proposed PCaLDI method are based on certain assumptions and constraints. The main limitation is the linear nature of PCA: PCA assumes linear relationships among data; therefore, when data have strong nonlinear relationships, PCA-based methods cannot adequately capture their structure. For example, when PCaLDI is applied to datasets with pronounced nonlinear relationships, such as gene expression or financial market data, the principal components may have limitations in capturing essential data features. However, the PCaLDI method is effective when the nonlinear relationships are weak, such as the text data that were considered in this study. In the case of nonlinear relationships, approximate inverse reactants can be employed using AIME [38]. AIME has already been demonstrated to resolve nonlinear relationships such as those in CNNs by means of golden-order inverse reactants.

Furthermore, the sample size should exceed the number of features for high-dimensional datasets. This is important to avoid overfitting and to extract principal components that reflect the true structure of the data. For example, in bioinformatics studies with thousands of features but limited sample sizes, dimensionality reduction and data filtering should be considered before PCA. This can also be achieved using AIME to derive the important features in advance and to calculate them [38].

A limitation of PCA loadings is that if a particular principal component is strongly dependent on a particular variable, the variation in that variable may substantially affect the interpretation of the principal component, which may also be the case with the PCaLDI method [59]. PCaLDI can include a PFI component to determine whether the feature is important.

Assessing the importance of a feature using PFI aids in understanding which variables best explain the variance in the data; however, it is computationally expensive and does not directly account for interactions between features. This limitation arises because the PCaLDI method uses PCA and limits the number of features to permutation from the PCA loading values, making it computationally less expensive than regular PFI.

C. COMPUTATIONAL CONSIDERATION

Fig. 2 demonstrates the overall framework of PCaLDI as formulated thus far. It consists of the normalize and PCA function, normalization and principal component mapping

function, top k important feature extraction function, permutation and evaluation function, and similarity or distance computation function.

The normalize and PCA function performs normalization and PCA using dataset X in the field as input, for which (1) to (3) in III-B are applicable. The component mapping function takes x_τ, x_v from the user as input and performs normalization and PCA using the normalized PCA function. The top k important feature extraction function uses PCA loadings to represent the relationship between the original features and principal components, x'_τ, x'_v. The similarity or distance computation function computes the distance between two vectors (in this case x'_τ, x'_v) using the PCA loadings that represent the relationship between the original features and principal components, which corresponds to (5) in III-B. The permutation and evaluation function performs permutations on the top k important features specified by the top k important feature extraction function, computes the difference between the similarity or distance computed by the similarity or distance computation function, and determines the features to be important if the difference is larger than the original similarity or distance. The permutation and evaluation function follows (9) to derive the features that contribute to the similarity or distance of the two vectors x'_τ, x'_v.
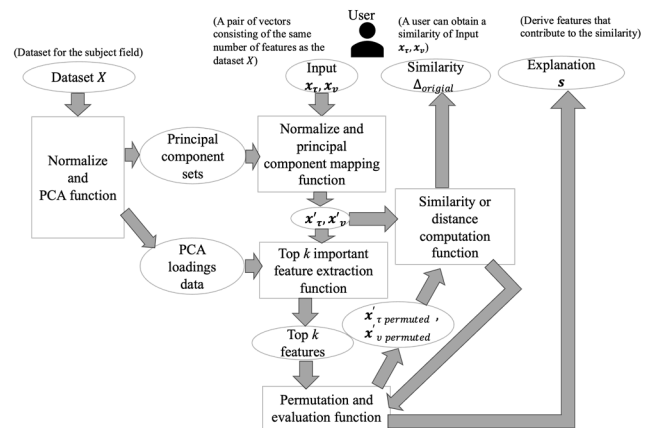

FIGURE 2. Flow of the PCaLDI algorithm.

Given the computational complexity of PCA, calculating the covariance matrix requires $O(m \times n^2)$ time, whereas eigenvalue decomposition requires $O(n^3)$ time. Consequently, the frequent recomputation of PCA, particularly when using PFI, is impractical for large datasets with numerous features. However, calculating the contribution score is computationally efficient and requires only $O(n)$ time for both the Hadamard and inner products. Therefore, an overall computational complexity of $O(n)$ is obtained. Thus, PCaLDI is cost effective, enabling initial contribution score calculations that can be refined to focus on the top k features of the PFI.

Thus, employing PCA requires retaining both the original dataset and PCA transformation matrix in memory. The transformation matrix is used to project the original dataset onto the principal component space. This is particularly

concerning when dealing with large datasets, in terms of both the volume and feature count, as it leads to increased memory usage.

Standardizing or normalizing the original dataset is recommended when using this method. Features with different scales can significantly skew PCA outcomes; in particular, features on larger scales can disproportionately influence the orientation of the principal components [60], their loadings, and even the PCaLDI contribution calculations.

Parallelization and distributed computing [61] can be implemented to facilitate rapid processing of large datasets. However, this approach introduces challenges such as ensuring equal data partitioning, effective integration of partial results, and calculation synchronization.

Several strategies can be adopted to optimize PCA calculations. For instance, Sparse PCA [62] assumes that the principal components for computing PCA efficiently are sparse. Randomized PCA [63] offers a rapid approximation method for large datasets. Incremental PCA [64] is memory efficient and processes data in smaller batches. Leveraging these methods in combination can make PCA computations highly efficient in various scenarios. The use of hardware that is optimized for parallel computing, such as graphics processing units, can further accelerate these calculations. This can pave the way for the application of PCaLDI to very large datasets.

IV. EXPERIMENTS

I conducted a comprehensive evaluation of the PCaLDI system, with a specific focus on explanations based on cosine similarity for both the tabular and textual data. The cosine similarity is calculated as the normalized dot product of two vectors, representing the cosine of the angle between them [65], and it measures how similar two vectors are in terms of their orientation. It is particularly useful for datasets in which directional information is significant. This consideration is pivotal when evaluating the similarity in a space that is defined solely by two vectors. Although the utility of various metrics may be considered for such an analysis, as geometrically evident, two vectors of the same magnitude will have a shorter Euclidean distance between them if their mutual angle is smaller, as can be observed with points on the circumference of a circle. Therefore, in this case, the experiment was limited to the cosine similarity.

The experimental setup and methodology are detailed in Section IV-A, which outlines the evaluation procedures and criteria. My initial experiment, described in Section IV-B, involved dataset X, which was designed with randomly assigned numbers and an incrementally increasing number of features. This experiment aimed to assess the computational efficiency of PCaLDI under varying conditions. Subsequently, Sections IV-C and IV-D present Experiments 2 and 3, which employed tabular and textual data, respectively, to further evaluate the applicability and robustness of the system in various data modalities.

A. EXPERIMENTAL ENVIRONMENT

I employed the Wine dataset [66] and Breast Cancer Wisconsin dataset [67] as tabular data and the 20 Newsgroups dataset [68] as text data. Both the Wine and Breast Cancer Wisconsin datasets were subjected to min–max normalization and were sourced from Sklearn datasets.

Three baseline methods were employed for the comparative analysis:

- All-feature permutation with PCA (AF-PCA): This approach applies PCA across all features in the dataset and then calculates the permutation importance for each feature after PCA. Although this method assumes that all features retain their importance after PCA, it is also computationally intensive.
- Variance-based selection with PCA (Var-PCA): This method calculates the variance of each feature in an original dataset. Features with high variance are prioritized, and PCA is applied only to these selected features. Thereafter, the permutation importance is calculated for this subset of features. This method assumes that features with high variance in the original data are important.
- All-feature permutation without PCA (AF-NoPCA): This approach calculates the permutation importance for all features in the original dataset without applying PCA. It operates under the assumption that no dimensionality reduction is required and considers all features important.

Although the baseline methods used (AF-PCA, Var-PCA, and AF-NoPCA) were adapted from standard practices for the application of PCA and PFI, they were specifically devised for this study to demonstrate their computational and methodological differences with respect to PCaLDI. These adaptations represent typical variations in the application of PCA and PFI, which are widely understood in the field, rather than novel methods that require specific references. Their inclusion highlighted the innovative approach of PCaLDI in terms of feature importance and computational efficiency.

Note that although the level of agreement between PCaLDI and these baseline methods may offer a measure of approximation, it does not necessarily attest to their accuracy. The low agreement rate does not preclude the possibility that PCaLDI can still effectively identify important features from a different analytical perspective. Hence, a qualitative evaluation was also included in this study by comparing specific output results.

In Experiment 3, I omitted the results for AF-PCA and AF-NoPCA because these computationally heavy baselines could not be calculated in real time. Future studies may address this limitation, potentially via parallel computation or other techniques, as discussed in Section III-D.

The experimental setup was implemented using Python 3.10.12 on Google Colab Pro, leveraging libraries such as scikit-learn 1.2.2, numpy 1.23.5, pandas 1.5.3, NLTK 3.8.1, and matplotlib 3.7.1.

B. EXPERIMENT 1: VERIFICATION OF COMPUTATIONAL COST USING RANDOM MATRICES WITH DIFFERENT NUMBERS OF FEATURES

In this experiment, I assessed the computational time required for different numbers of features using the proposed PCaLDI and baseline methods. Specifically, I set the data size to 1000 records and generated random number matrices with varying numbers of features: 10, 50, 100, 150, 200, 250, 300, 350, 400, 450, and 500. The average computation time for deriving the feature contributions was measured over 10 runs, with $k = 10$, using the cosine similarity as the metric.

The results illustrated in Fig. 3 show that the computation time for AF-PCA increased sharply and linearly as the number of features increased. Although PCA is effective in mitigating issues such as dimensionality, information redundancy, and multicollinearity, its computational demands make calculating the PFI impractical as the feature count increases. By contrast, the other methods, including PCaLDI, exhibited only a moderate increase in the computation time when the number of features increased, making them more practical for real-world applications. The influence of the number of principal components on the computational efficiency of PCaLDI was further investigated in Experiment 2. Interestingly, the computational time for AF-NoPCA remained relatively stable and was not substantially burdensome, even without the incorporation of PCA or ML retraining. The observed stability in the computation time for the AF-NoPCA method can be attributed to the simplicity of the permutation importance computation, which does not involve complexity owing to dimensionality reduction or model retraining. This approach directly evaluates the importance of features in the original space, thereby avoiding the computational overhead associated with PCA transformations. However, the scenario changes when the method is applied to TF-IDF matrices, which inherently have a high-dimensional feature space because of the vast number of unique words that are treated as individual features. In such cases, the computational demand increases significantly because permutation importance must be computed for each word across all documents, leading to a drastic increase in the number of required calculations.

In summary, although I had to cap the number of features to 500 for AF-PCA owing to computational constraints, the experiment provided valuable insights into the computational efficiency of each method. Importantly, the results demonstrate that PCaLDI can derive feature contributions at lower computational costs, making it a practical choice for real-world applications.

C. EXPERIMENT 2: EFFECTIVENESS OF PCaLDI FOR TABULAR DATA

1) QUANTITATIVE EVALUATION OF PCaLDI USING TABULAR DATA

In this experiment, I used two distinct datasets: Wine [66] and Breast Cancer Wisconsin [67]. The goal was to investigate the commonalities and divergences in trends between the two datasets. I investigated the results to highlight the key features extracted by each method.

I assessed the shifts in the number of principal components when comparing the proposed PCaLDI method with baseline approaches. I set $k = 5$ for the Wine dataset and $k = 10$ for the Breast Cancer Wisconsin dataset. The correlations in the number of principal components are illustrated in Figs. 4 and 5 for the Wine and Breast Cancer Wisconsin datasets, respectively. Note that these results represent the average of 100 calculations each, as the outcomes may be unstable for smaller datasets owing to the inherent randomness in calculating the PFI. Here, the "degree of agreement" refers to the inclusion of both feature names in the specified top k features. The cosine similarity metric was used in this experiment.
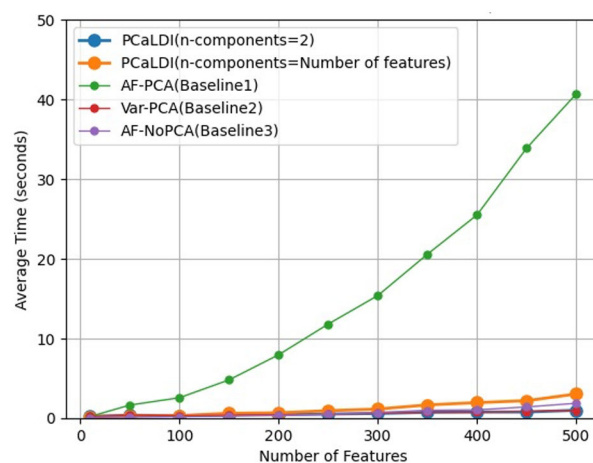

FIGURE 3. Computation time with increasing number of features for the proposed PCaLDI method and baseline methods AF-PCA, Var-PCA, and AF-NoPCA.

For the Wine dataset, as depicted in Fig. 4, the highest level of concordance between the PCaLDI method and AF-PCA (Baseline1) occurred while using four principal components. This result suggests that the principal variations within the dataset are primarily captured by the first four components and that adding more components may introduce noise or irrelevant information. Although some minor agreement with other baselines was observed for one and two principal components, relative consistency was observed beyond that, indicating that the number of principal components was not a significant variable in this context.

For the Breast Cancer Wisconsin dataset, as illustrated in Fig. 5, the agreement between PCaLDI and AF-PCA (Baseline1) peaked when four or five principal components were used. This reiterates that major variances are captured effectively by these components and that adding more components may introduce noise. For the other baseline methods, a peak in agreement occurred at the sixth principal component; lesser stability and plateauing were observed when fewer and more than six components, respectively, were used.
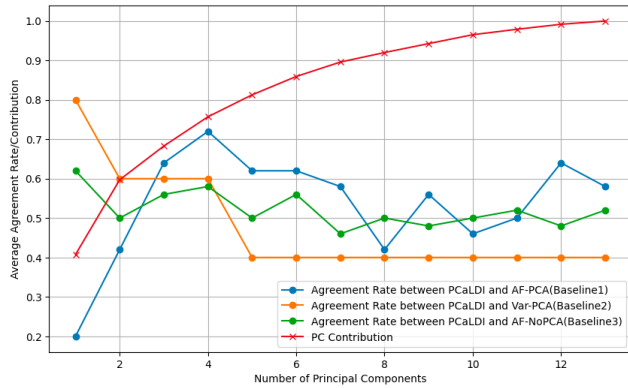
FIGURE 4. Agreement between the proposed method and baselines with varying numbers of principal components for the Wine dataset.
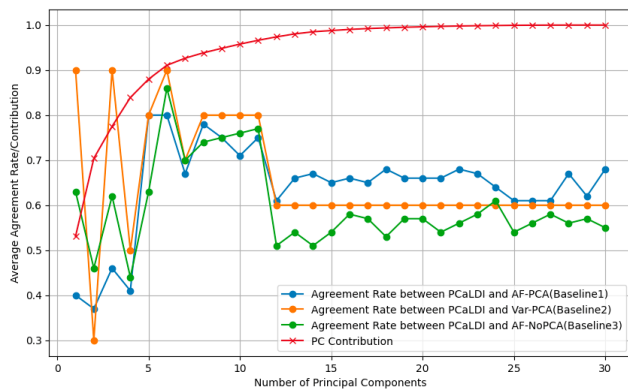

FIGURE 5. Agreement between the proposed method and baselines with varying numbers of principal components for the Breast Cancer Wisconsin dataset.

This pattern suggests that the agreement rates between the proposed PCaLDI and AF-PCA (Baseline1) were optimized for a specific number of principal components. Beyond this number, the agreement rates remained steady. By contrast, the agreement rates for the other baselines were less influenced by the number of principal components. This finding suggests that these methods may be insensitive to PCA results or may not be significantly affected by them in terms of feature selection. The detected trend, whereby the concordance between the novel PCaLDI method and AF-PCA baseline stabilized after a particular count of principal components was introduced, aligns with the conventional wisdom in the realm of feature reduction techniques. According to Jolliffe [60], including principal components past a certain threshold generally results in a trivial enhancement in the variance. This enhancement is often insufficient for significantly improving the performance or comprehensibility of a model. Consistent with the observations made in the present study, this phenomenon implies the presence of an optimal count of principal components for effective feature selection. This theme of diminishing marginal utility with an increasing number of components is a recognized aspect of

diverse PCA applications, and it suggests an inherent trait of PCA-centric feature selection approaches. For instance, Var-PCA (Baseline2) may not have been influenced by the number of principal components because it selects features based on high variance, regardless of PCA. Similarly, NoPCA (Baseline3) has no inherent connection to PCA. The decline in agreement rates for PCaLDI and AF-PCA after the peak suggests that additional principal components could make the data either redundant or more noise prone.

Fig. 6 presents the agreement rates between PCaLDI and the individual baseline methods (AF-PCA, Var-PCA, and AF-NoPCA) when the number of principal components was set to three and $k = 5$ for the Wine dataset and to three and $k = 10$ for the Breast Cancer Wisconsin dataset. Overall, the agreement was higher for the Wine dataset and slightly lower for the Breast Cancer Wisconsin dataset. This result indicates that Var-PCA aligns most closely with PCaLDI in terms of feature selection for the Wisconsin Breast Cancer dataset. In the case of the Wine dataset, PCaLDI serves as a reasonable approximation to other methods, probably because the Wine dataset is simpler than the Breast Cancer Wisconsin dataset, making it easier to capture its diversity and complexity.

2) QUALITATIVE EVALUATION OF PCaLDI USING FOCUSED DATA
Next, I analyzed the actual feature contributions for specific data pairs. For the Breast Cancer Wisconsin dataset, I focused on pairs with the highest cosine similarity and three randomly selected pairs. In my tests, the pair with the highest cosine similarity was (174, 304) and the randomly selected pairs were (308, 81), (512, 527), and (243, 56). The cosine similarities for the target data pairs are presented in Table 1.
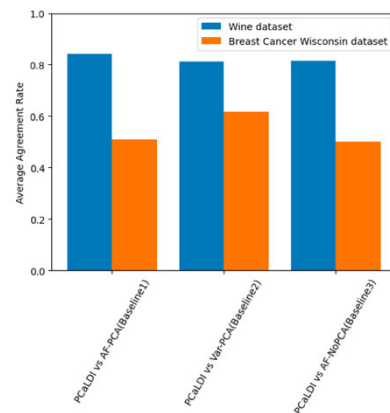

FIGURE 6. Agreement between the proposed method and baselines for the Wine and Breast Cancer Wisconsin datasets.

Figs. 7, 8, 9, and 10 show the output contributions for the pair with the highest cosine similarities (174, 304), as computed via the proposed PCaLDI method and the baseline methods AF-PCA, Var-PCA, and AF-NoPCA, respectively. Fig. 11 further highlights the agreement between these methods. As shown in Figs. 7, 8, and 9, all three PCA-based

methods, namely PCaLDI, AF-PCA, and Var-PCA, identified "worst concave points" as a major contributor to the cosine similarity. The contribution of this feature eclipses that of the other features, implying that it largely dictates the cosine similarity for a specific pair of data instances. In stark contrast, the AF-NoPCA output in Fig. 10 shows that "mean symmetry" had the largest contribution. This discrepancy suggests that AF-NoPCA, which operates without PCA, considers a broader array of features, thereby capturing different facets of data. AF-PCA also outputs "worst concave point" at the top. This could be because PCA inherently magnifies the significance of this feature. Conversely, AF-NoPCA, which does not employ PCA, potentially captures a broader range of feature diversity and places a greater emphasis on alternative feature combinations. In general, these results show a clear distinction between the methods that employ PCA and those that do not. Methods using PCA highlight "worst concave points" as a key feature contributing to data-pair similarity, whereas non-PCA methods consider a more diversified set of features. According to Fig. 11, the proposed PCaLDI, AF-PCA, and Var-PCA methods all showed similar trends, except for the case of "worst concave points," which exhibited lower values, thereby resulting in minor deviations in the agreement levels. However, the overarching directionality remained consistent across the methods.

TABLE 1. Cosine similarity values for pairs of target data instances.

Target pair of data instances	Cosine similarity value
174, 304 (highest cosine similarity)	0.999992
308, 81 (random selection)	-0.722545
512, 527 (random selection)	-0.727053
243, 56 (random selection)	-0.524537
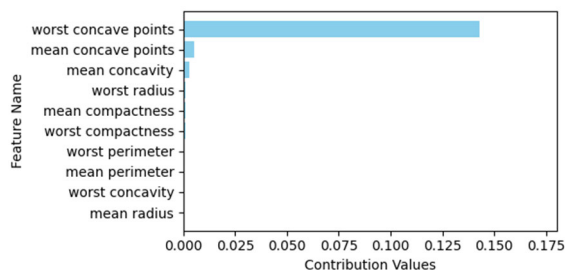
FIGURE 7. Feature derivation results contributing to the cosine similarity of data instance pair (174, 304) using PCaLDI.

Figs. 12, 13, 14, and 15 show the feature contributions determined by the proposed PCaLDI method and baseline methods AF-PCA, Var-PCA, and AF-NoPCA, respectively, for the randomly selected data pair (308, 81). The degree of agreement between the methods is shown in Fig. 16. In line with earlier observations, "worst concave points" emerged as the dominant feature influencing the similarity for PCaLDI and AF-NoPCA in Figs. 12 and 14. This result confirms that
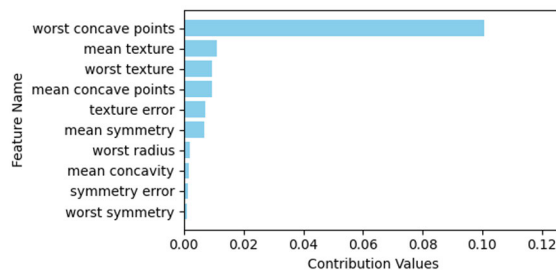
FIGURE 8. Feature derivation results contributing to the cosine similarity of data instance pair (174, 304) using AF-PCA.
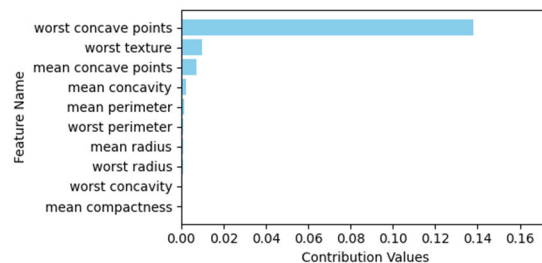
FIGURE 9. Feature derivation results contributing to the cosine similarity of the data instance pair (174, 304) using Var-PCA.
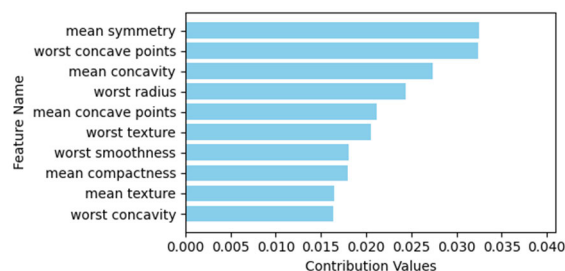

FIGURE 10. Feature derivation results contributing to the cosine similarity of data instance pair (174, 304) using AF-NoPCA.
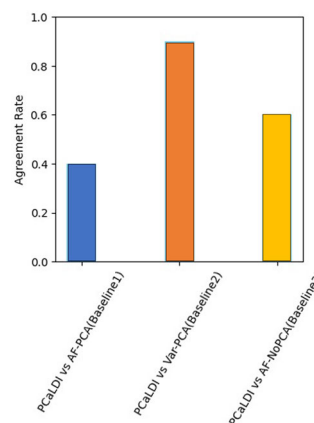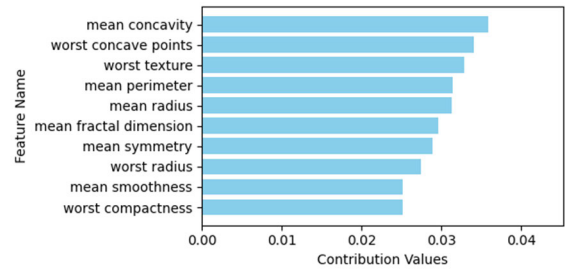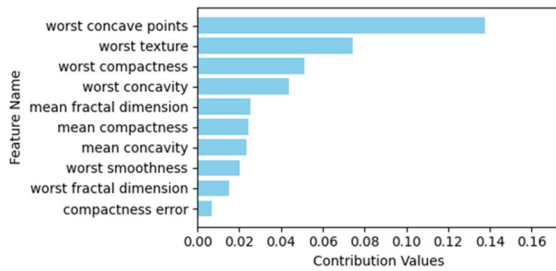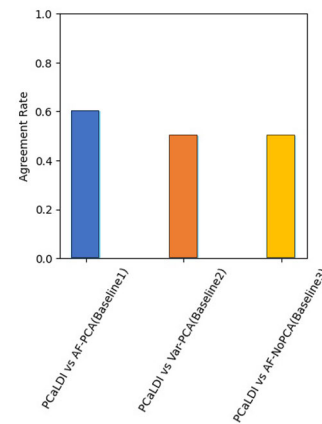

FIGURE 11. Degree of agreement between the proposed PCaLDI and baseline methods AF-PCA, Var-PCA, and AF-NoPCA contributing to the cosine similarity of data instance pair (174, 304).
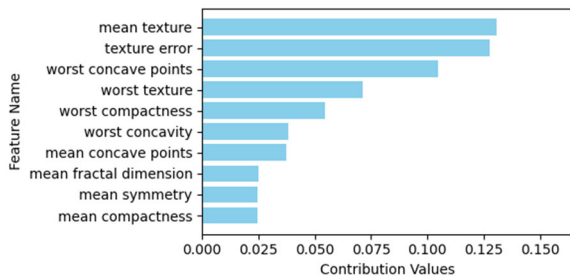
this feature plays a pivotal role in a dataset. Fig. 13 shows that AF-PCA elevated "mean texture" as the most consequential feature. In addition, both "texture error" and "worst

concave points'' featured prominently in the rankings. For AF-NoPCA in Fig. 15, ''mean concavity,'' ''worst concave point,'' and ''worst text'' were the top features. The agreement values depicted in Fig. 16 (0.6 for AF-PCA and 0.5 each for Var-PCA and AF-NoPCA) suggest that texture-related features significantly influence the similarity metrics for this specific pair of data points. Notably, these agreement values were higher than those observed in previous examples for features not topping the list, hinting at a more diverse set of influential features as shown in Figs. 12, 13, 14, and 15 compared with Figs. 7, 8, 9, and 10.

FIGURE 15. Feature derivation results contributing to the cosine similarity of data instance pair (308, 81) using AF-NoPCA.

FIGURE 12. Feature derivation results contributing to the cosine similarity of data instance pair (308, 81) using PCaLDI.

FIGURE 13. Feature derivation results contributing to the cosine similarity of data instance pair (308, 81) using AF-PCA.
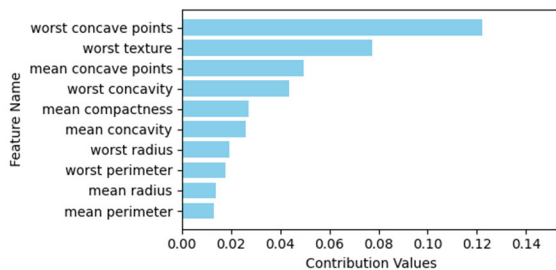
FIGURE 14. Feature derivation results contributing to the cosine similarity of data instance pair (308, 81) using Var-PCA.

Figs. 17, 18, 19, and 20 present the feature contributions calculated using the proposed PCaLDI and baseline methods AF-PCA, Var-PCA, and AF-NoPCA, respectively, for the randomly selected data pair (512, 527). The corresponding agreement metrics for these methods are shown in Fig. 21. According to Figs. 17, 16, and 19, ''worst texture'' was identified as a highly influential feature with values of 0.369384 for

FIGURE 16. Degree of agreement between the proposed PCaLDI and baselines AF-PCA, Var-PCA, and AF-NoPCA contributing to the cosine similarity of data instance pair (308, 81).

PCaLDI, 0.370127 for AF-PCA, and 0.375162 for Var-PCA. In addition, both ''worst concave points'' and ''mean concave points'' were ranked among the top features, confirming their significance for this data pair. However, Fig. 20 reveals that, in AF-NoPCA, ''worst concave points'' was the highest contributing feature. Note that the lower values of ''worst concave points'' in AF-NoPCA than in the other methods may be attributed to the absence of dimensionality reduction in AF-NoPCA. Without PCA-based feature extraction, AF-NoPCA evaluates the importance of features in the original space, where the distribution and scale of each feature remain unchanged. This may result in different features being identified as more influential, owing to the preservation of the original data structure, which may not be the case in PCA-based methods that realign the feature space. As shown in Fig. 21, the agreement levels were 0.7 for PCaLDI and AD-PCA, 0.6 for PCaLDI and Var-PCA, and 0.4 for PCaLDI and AF-NoPCA. These levels highlight the strong influence of ''worst texture'' on this data pair, especially in methods that employ PCA. This result suggests that PCA tends to amplify the impact of specific features. By contrast, AF-NoPCA, which does not utilize PCA, appears to diversify its focus, possibly by assigning greater weights to different aspects or combinations of features. This disparity between

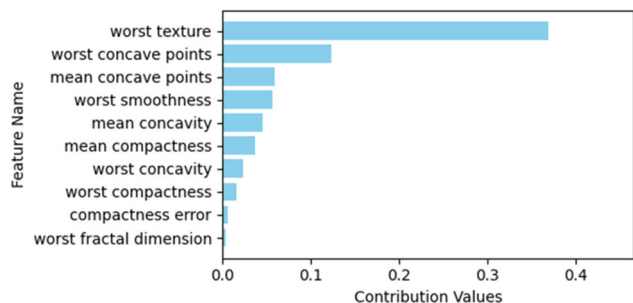the PCA and non-PCA methods emphasizes the tendency of PCA to prioritize certain features over others.

FIGURE 17. Feature derivation results contributing to the cosine similarity of data instance pair (512, 527) using PCaLDI.
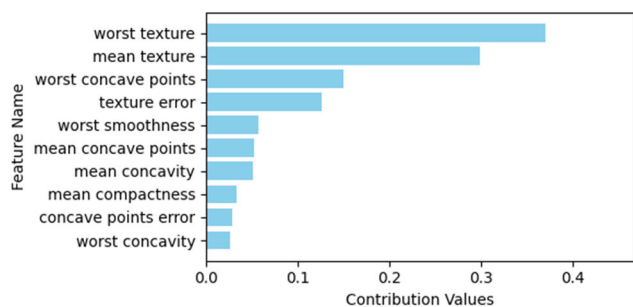

FIGURE 18. Feature derivation results contributing to the cosine similarity of data instance pair (512, 527) using AF-PCA.
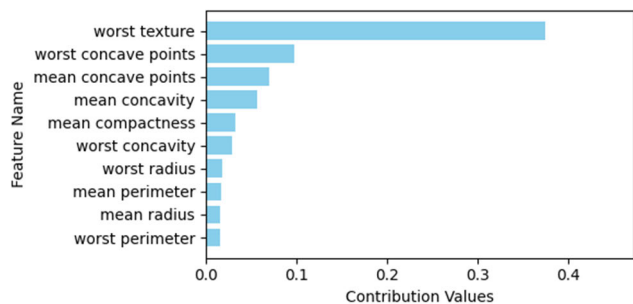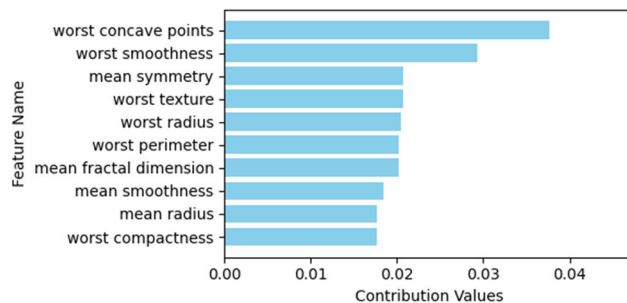

FIGURE 19. Feature derivation results contributing to the cosine similarity of data instance pair (512, 527) using Var-PCA.

Figs. 22, 23, 24, and 25 present the feature contributions from four different methods: the proposed PCaLDI and baseline AF-PCA, Var-PCA, and AF-NoPCA methods, respectively. These figures depict the contributions of the randomly selected data pair (243, 56). The agreement between these methods is shown in Fig. 26. According to Figs. 22, 23, 24, and 25, "worst concave points" emerged as the most influential feature for PCaLDI, AF-PCA, Var-PCA, and AF-NoPCA, respectively. Therefore, this feature likely drives the majority of the cosine similarity for this data pair. In addition, "mean fractal dimension" was ranked as the second most influential feature in PCaLDI and AF-PCA,

FIGURE 20. Feature derivation results contributing to the cosine similarity of data instance pair (512, 527) using AF-NoPCA.
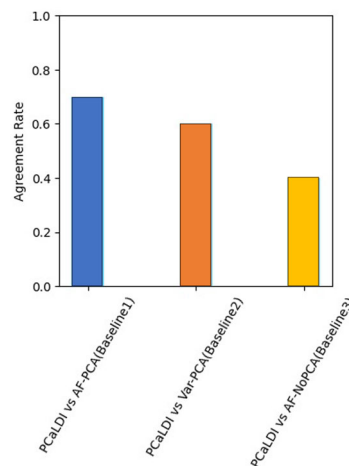

FIGURE 21. Degree of agreement between the proposed PCaLDI and baselines AF-PCA, Var-PCA, and AF-NoPCA contributing to the cosine similarity of data instance pair (512, 527).

as shown in Figs. 22 and 23, respectively. Interestingly, this feature did not appear in the top 10 features of Var-PCA. However, "mean concave points" consistently ranked in the top 10 across PCaLDI, AF-PCA, and Var-PCA, confirming its importance. By contrast, the AF-NoPCA data in Fig. 25 deviate significantly from the observed trends of the other methods. In AF-NoPCA, "worst smoothness" was ranked eighth, whereas it did not feature in the top 10 in the other methods. This result, combined with the generally lower values observed in AF-NoPCA, suggests that this method has a poorer discriminative ability than those of the other methods. In terms of the agreement shown in Fig. 26, the concordance values for PCaLDI and AF-PCA, Var-PCA, and AF-NoPCA were 0.4, 0.8, and 0.5, respectively.

Figs. 27, 28, 29, and 30 present the feature outputs from the proposed PCaLDI and baseline methods AF-PCA, Var-PCA, and AF-NoPCA, respectively, for the data pair (174, 304), which exhibited the highest cosine similarity. The agreement between these methods is shown in Fig. 31. "Worst concave points" was prominent as the main contributor to the cosine similarity in the PCA methods (PCaLDI, AF-PCA, and Var-PCA), as shown in Figs. 27, 28, and 29. According to Fig. 30, only AF-NoPCA ranked "mean concavity" first owing to

the absence of PCA in the AF-NoPCA method. Regarding the agreement depicted in Fig. 31, Var-PCA exhibited a higher concordance rate. However, when examined individually, PCaLDI, AF-PCA, and Var-PCA demonstrated similar directional tendencies, whereas AF-NoPCA exhibited a slight variation.
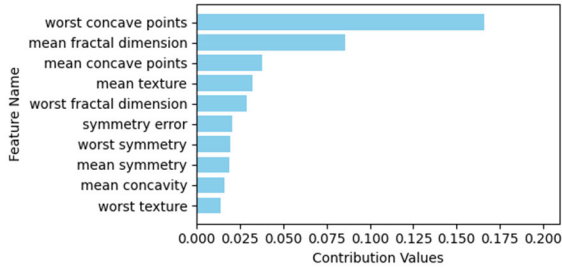

FIGURE 22. Feature derivation results contributing to the cosine similarity of data instance pair (243, 56) using PCaLDI.
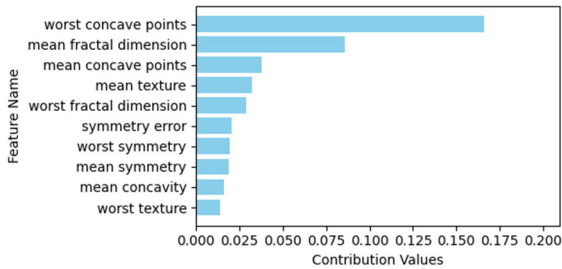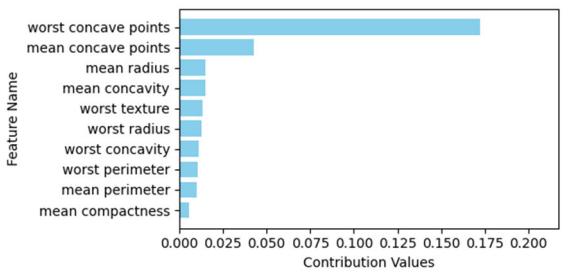

FIGURE 23. Feature derivation results contributing to the cosine similarity of data instance pair (243, 56) using AF-PCA.

FIGURE 24. Feature derivation results contributing to the cosine similarity of data instance pair (243, 56) using Var-PCA.

FIGURE 25. Feature derivation results contributing to the cosine similarity of data instance pair (243, 56) using AF-NoPCA.

FIGURE 26. Degree of agreement between the proposed PCaLDI and baselines AF-PCA, Var-PCA, and AF-NoPCA contributing to the cosine similarity of data instance pair (243, 56).
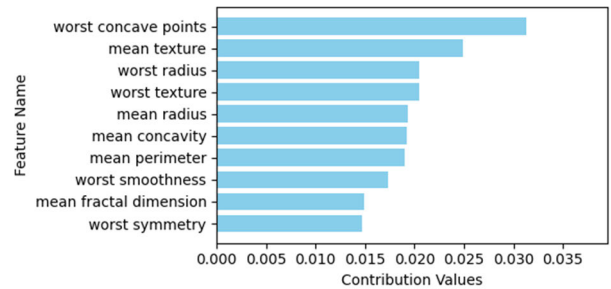
FIGURE 27. Feature derivation results contributing to the cosine similarity of data instance pair (174, 304) using PCaLDI.

The findings indicate that the proposed PCaLDI approach shares similarities with the baseline methods AF-PCA and Var-PCA, whereas AF-NoPCA diverges in its recommendations. Employing PFI across all features is ideal, as in AF-PCA and AF-NoPCA, to ascertain the precise feature contributions. However, PCaLDI offers a more computationally efficient alternative by leveraging PCA loadings and delivering results that approximate those of its computationally intensive counterparts.

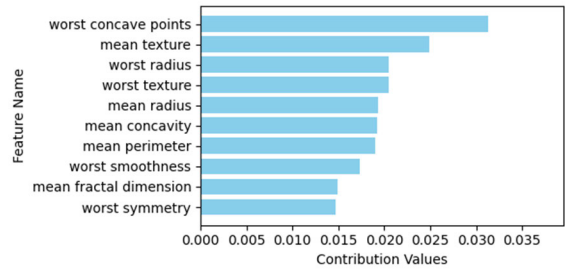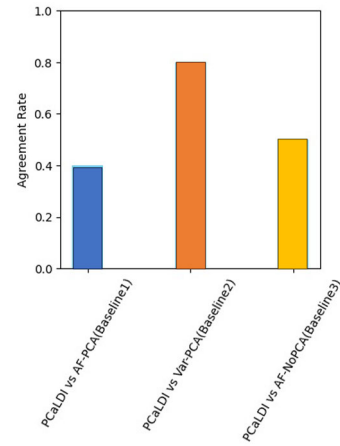In scenarios with a high feature count, computationally expensive methods such as AF-PCA and AF-NoPCA may become impractical. Under such circumstances, opting for PCaLDI or Var-PCA, both of which offer computational advantages, may be a pragmatic choice.

However, Var-PCA calculates the PFI only for its top k features that are predetermined by their variance. Given that the feature variance remains static, Var-PCA is likely to identify the same features as important repetitively depending on the data type. From this perspective, PCaLDI provides a more flexible and nuanced evaluation.
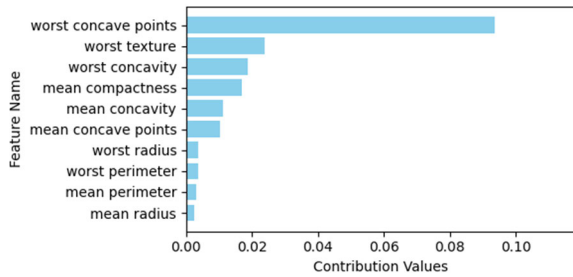
FIGURE 28. Feature derivation results contributing to the cosine similarity of data instance pair (174, 304) using AF-PCA.
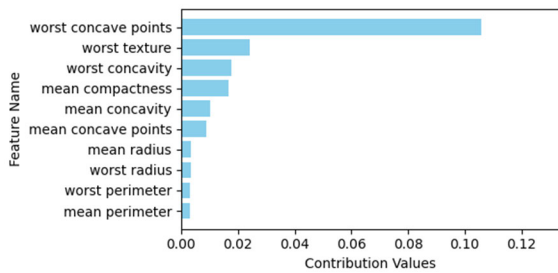

FIGURE 29. Feature derivation results contributing to the cosine similarity of data instance pair (174, 304) using Var-PCA.
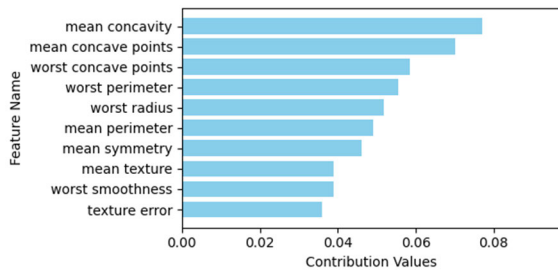

FIGURE 30. Feature derivation results contributing to the cosine similarity of data instance pair (174, 304) using AF-NoPCA.

D. EXPERIMENT 3: EFFECTIVENESS OF PCaLDI FOR TEXT DATA

Data from the 20 Newsgroups dataset [68] were employed in Experiment 3 to investigate the utility of PCaLDI in generating explanations for text data. Specifically, I used posts from the "rec.sport.baseball" and "rec.sport.hockey" categories. Posts with only spaces, line breaks, or fewer than 15 words were excluded, leaving 953 records for analysis. The generated TF-IDF matrix for this dataset had dimensions of 953×12733. Subsequently, I applied data reduction to a 100-dimensional principal component space using 200 principal components. The data instance pair selected for this experiment was that with the highest cosine similarity (569, 716).

Fig. 32 displays the top 20 features (words) contributing to the cosine similarity for data pair (569, 716). Fig. 32 reveals that common words such as "go," "york," and "let" ranked highly. In addition, sport-specific terms such as "ranger," "game," and "champ" confirmed that these texts
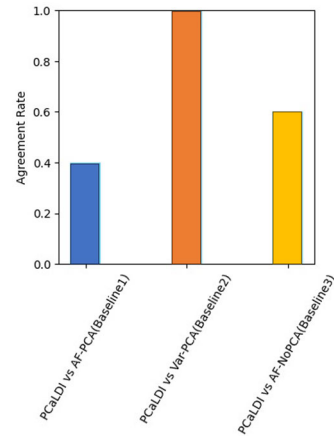

FIGURE 31. Degree of agreement between the proposed PCaLDI and baselines AF-PCA, Var-PCA, and AF-NoPCA contributing to the cosine similarity of data instance pair (174, 304).

are sports oriented, particularly concerning field hockey. The term "ranger" appears to refer to a hockey team. Academia-related words such as "upenn," "kkeller," "quaker," and "ivy" also featured prominently, suggesting that both texts might focus on themes connecting sports and educational institutions such as the University of Pennsylvania and the Ivy League.
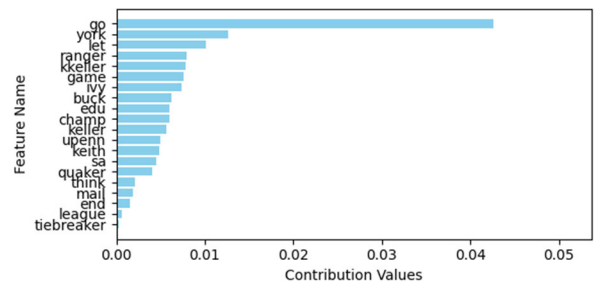

FIGURE 32. Feature derivation results contributing to the cosine similarity of data instance pair (569, 716) using PCaLDI.

These experimental results demonstrate that the proposed PCaLDI method is useful for extracting meaningful features or words that elucidate why two data instances exhibit high cosine similarity. The process begins with vectorizing the text data using methods such as TF-IDF, followed by the application of PCA for dimensionality reduction. This paves the way for explaining the measures of cosine similarity without requiring the computation of the PFI for each feature or word individually. Consequently, PCaLDI provides a unique, approximate perspective for identifying contributing features while reducing the computational overhead substantially. In the experimental setting of the present study, the average computation time for PCaLDI was approximately 90.97 s, with a standard deviation of 1.163 for five data instance pairs. Given that the initial 953×12733 matrix was compressed to 200 dimensions using PCA and that salient features were identified from among 12733 possibilities, this computation time is within practical limits.

V. DISCUSSION

PCaLDI operates by using the loadings and values within the principal component space for the data instances under scrutiny. This aids in pinpointing the top k features that contribute to cosine similarity. Subsequently, the PFI is calculated using only the top k features. This drastically reduces the computational complexity and yields a nuanced yet accurate perspective of the feature importance. The efficacy of PCaLDI was rigorously validated against the three baseline methods used in this study.

Moreover, PCaLDI addresses the "curse of dimensionality," which is a phenomenon that is encountered in high-dimensional data spaces that can diminish the discernible distance between data points. Consequently, the similarity and distance calculations are affected. PCaLDI overcomes this challenge by employing PCA for dimensionality reduction, thereby enabling a more accurate evaluation of feature contributions.

The findings of Experiment 3 on the text data are particularly compelling. Salient features or words were successfully distilled from a high-dimensional feature space, which significantly contributed to the cosine similarity between data instance pairs. These results suggest the potential of PCaLDI as a novel text-mining tool.

However, this approach has some limitations. PCA tends to falter with nonlinear data structures, noisy datasets, or datasets that are replete with outliers. Although other dimensionality reduction techniques such as t-SNE [57] and UMAP [58] are well suited for such data, they were not investigated in this study, leaving room for future enhancements.

Future studies should consider integrating nonlinear dimensionality reduction methods by introducing kernel PCA, autoencoders, and AIME from previous studies to improve the applicability of PCaLDI to nonlinear datasets. I believe that this will enable us to understand the data structure more accurately and gain useful insights, even in complex datasets with nonlinear relationships.

In addition, this study focused on cosine similarity metrics to demonstrate the applicability of PCaLDI in various scenarios. However, this framework can be extended to incorporate other vector evaluation methods, thereby providing an avenue for future research. For example, investigating the use of PCaLDI with metrics such as the Manhattan distance [65], which measures the absolute differences between points in a vector space, or the Minkowski distance [69], which is a flexible metric that can be adapted to different scales, could reveal further insights. In addition, the Jaccard index [65], which assesses the similarity based on members that are shared between sets, could provide an alternative perspective when analyzing categorical data. These potential extensions highlight the wide applicability of the PCaLDI framework and its relevance to various analytical challenges. In future, this framework can be enhanced by integrating more advanced vector evaluation methods that are beyond the scope of PCA. Techniques such as t-SNE [57] and

UMAP [58], which lack the explicit concept of an explained variance ratio, as in PCA, can benefit from the application of AIME [38] to compute their embeddings in an inverse manner. The combination of PCaLDI and AIME [38] could pave the way for innovative explanations of the complex nonlinear relationships inherent to the reduced-dimensional spaces created by these methods. These advancements will enhance the interpretative power of PCaLDI and expand its utility in uncovering the intricate structure of data in various analytical scenarios.

This method for deriving explanations for cosine similarity also sets the stage for creating interpretability frameworks for unsupervised ML algorithms that leverage these measures, such as k-means clustering. For example, PCaLDI can be applied to explain why a particular set of data points belongs to a specific cluster and to identify and compare the feature contributions of the centroid of each cluster. This is a compelling direction for future research.

Finally, although PCaLDI reduces computational costs, both computation time and memory usage increase with larger datasets and more features. Further research should address the issues surrounding the selection of computational resources and implementation of parallel and distributed processing to overcome these challenges.

In conclusion, the overarching aim of this study was to develop explainable metrics, for which PCaLDI serves as the foundational step. This method may be further refined and extended in future studies.

VI. CONCLUSION

In response to the challenges of analyzing high-dimensional data and understanding the factors that contribute to data similarity and distance, this study has introduced PCaLDI. As datasets become increasingly dimensional, traditional methods cannot provide both explanatory power and high computational efficiency. PCaLDI offers a solution that strikes a balance between these aspects, thereby providing a new perspective on high-dimensional data.

PCaLDI is an efficient method for identifying features that influence the similarity and distance metrics in high-dimensional datasets. This approach not only aids in uncovering the mechanisms behind the similarity and distance but also offers the potential to deepen our understanding of data and pave the way for further research. PCaLDI has been demonstrated as particularly effective in cases where the original data have numerous features, such as text data. In addition, PCaLDI addresses the computational inefficiencies of existing methods by isolating the most relevant features for data similarity and distance.

The effectiveness of PCaLDI was rigorously assessed and demonstrated via quantitative and qualitative comparisons with three baseline methods. The applicability of PCaLDI was established using a well-known text dataset to extract words that contribute to the cosine similarity. The results confirmed that PCaLDI could uncover new contributing features with an improved computational efficiency.

Similar to other analytical methods, PCaLDI is influenced by the intrinsic properties of the data that it processes. It may not be suitable for datasets with nonlinear structures, significant noise, or outliers. Future studies could investigate the integration of other dimensionality reduction techniques to overcome these limitations. Furthermore, PCaLDI offers potential beyond the scope of this study, with opportunities to adapt and refine the method for various applications in fields that require XAI.

In summary, PCaLDI serves as a robust tool for enhancing transparent and reliable decision-making in the era of big data. This is especially important given the growing demand for XAI and ML models in critical sectors such as healthcare, finance, and public policy. Thus, this study represents a significant advancement toward a more comprehensive and nuanced understanding of data similarity and distance.

ACKNOWLEDGMENT

REFERENCES

[1] P. C. Sen, M. Hajra, and M. Ghosh, "Supervised classification algorithms in machine learning: A survey and review," in *Advances in Intelligent Systems and Computing*. Cham, Switzerland: Springer, 2020, pp. 99–111.

[2] B. Mahesh, "Machine learning algorithms—A review," *Int. J. Sci. Res.*, vol. 9, pp. 381–386, Jan. 2020.

[3] E. Cambria, L. Malandri, F. Mercorio, M. Mezzanzanica, and N. Nobani, "A survey on XAI and natural language explanations," *Inf. Process. Manage.*, vol. 60, no. 1, Jan. 2023, Art. no. 103111, doi: 10.1016/j.ipm.2022.103111.

[4] W. Saeed and C. Omlin, "Explainable AI (XAI): A systematic meta-survey of current challenges and future opportunities," *Knowl.-Based Syst.*, vol. 263, Mar. 2023, Art. no. 110273, doi: 10.1016/j.knosys.2023.110273.

[5] Y.-N. Chuang, G. Wang, F. Yang, Z. Liu, X. Cai, M. Du, and X. Hu, "Efficient XAI techniques: A taxonomic survey," 2023, *arXiv:2302.03225*.

[6] G. Schwalbe and B. Finzel, "A comprehensive taxonomy for explainable artificial intelligence: A systematic survey of surveys on methods and concepts," *Data Mining Knowl. Discovery*, pp. 1–59, Jan. 2023. [Online]. Available: https://link.springer.com/article/10.1007/s10618-022-00867-8#citeas, doi: 10.1007/s10618-022-00867-8.

[7] R. Dazeley, P. Vamplew, and F. Cruz, "Explainable reinforcement learning for broad-XAI: A conceptual framework and survey," *Neural Comput. Appl.*, vol. 35, no. 23, pp. 16893–16916, Aug. 2023, doi: 10.1007/s00521-023-08423-1.

[8] W. Yang, Y. Wei, H. Wei, Y. Chen, G. Huang, X. Li, R. Li, N. Yao, X. Wang, X. Gu, M. B. Amin, and B. Kang, "Survey on explainable AI: From approaches, limitations and applications aspects," *Hum.-Centric Intell. Syst.*, vol. 3, no. 3, pp. 161–188, Aug. 2023, doi: 10.1007/s44230-023-00038-y.

[9] S. Ali, T. Abuhmed, S. El-Sappagh, K. Muhammad, J. M. Alonso-Moral, R. Confalonieri, R. Guidotti, J. Del Ser, N. Díaz-Rodríguez, and F. Herrera, "Explainable artificial intelligence (XAI): What we know and what is left to attain trustworthy artificial intelligence," *Inf. Fusion*, vol. 99, Nov. 2023, Art. no. 101805, doi: 10.1016/j.inffus.2023.101805.

[10] V. Chamola, V. Hassija, A. R. Sulthana, D. Ghosh, D. Dhingra, and B. Sikdar, "A review of trustworthy and explainable artificial intelligence (XAI)," *IEEE Access*, vol. 11, pp. 78994–79015, 2023, doi: 10.1109/ACCESS.2023.3294569.

[11] R. Dwivedi, D. Dave, H. Naik, S. Singhal, R. Omer, P. Patel, B. Qian, Z. Wen, T. Shah, G. Morgan, and R. Ranjan, "Explainable AI (XAI): Core ideas, techniques, and solutions," *ACM Comput. Surv.*, vol. 55, no. 9, pp. 1–33, Sep. 2023, doi: 10.1145/3561048.

[12] B. Boehmke and B. Greenwell, "Interpretable machine learning," in *Hands-On Machine Learning With R*. Morrisville, NC, USA: Lulu Com, 2019, pp. 305–342.

[13] D. Minh, H. X. Wang, Y. F. Li, and T. N. Nguyen, "Explainable artificial intelligence: A comprehensive review," *Artif. Intell. Rev.*, vol. 55, no. 5, pp. 3503–3568, Jun. 2022, doi: 10.1007/s10462-021-10088-y.

[14] A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence (XAI)," *IEEE Access*, vol. 6, pp. 52138–52160, 2018, doi: 10.1109/ACCESS.2018.2870052.

[15] Q. Zhang, Y. N. Wu, and S.-C. Zhu, "Interpretable convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8827–8836, doi: 10.1109/CVPR.2018.00920.

[16] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi, "A survey of methods for explaining black box models," *ACM Comput. Surv.*, vol. 51, no. 5, pp. 1–42, Sep. 2019, doi: 10.1145/3236009.

[17] D. V. Carvalho, E. M. Pereira, and J. S. Cardoso, "Machine learning interpretability: A survey on methods and metrics," *Electronics*, vol. 8, no. 8, p. 832, Jul. 2019, doi: 10.3390/electronics8080832.

[18] A. Barredo Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera, "Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Inf. Fusion*, vol. 58, pp. 82–115, Jun. 2020, doi: 10.1016/j.inffus.2019.12.012.

[19] T. Speith, "A review of taxonomies of explainable artificial intelligence (XAI) methods," in *Proc. ACM Conf. Fairness, Accountability, Transparency*. New York, NY, USA: Association for Computing Machinery, Jun. 2022, pp. 2239–2250, doi: 10.1145/3531146.3534639.

[20] W. Samek and K. R. Müller, "Towards explainable artificial intelligence," *Explainable AI, Interpret., Explain. Visual. Deep Learn.*, vol. 11700, pp. 5–22, 2019. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-030-28954-6_1

[21] M. Van Lent, W. Fisher, and M. Mancuso, "An explainable artificial intelligence system for small-unit tactical behavior," in *Proc. Nat. Conf. Artif. Intell.*, 2004, pp. 900–907.

[22] D. Gunning and D. Aha, "DARPA's explainable artificial intelligence (XAI) program," *AI Mag.*, vol. 40, no. 2, pp. 44–58, Jun. 2019.

[23] M. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?' Explaining the predictions of any classifier," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Demonstrations*, 2016, pp. 97–101, doi: 10.18653/v1/n16-3020.

[24] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–17.

[25] S. Bach, A. Binder, G. Montavon, F. Klauschen, K.-R. Müller, and W. Samek, "On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation," *PLoS ONE*, vol. 10, no. 7, Jul. 2015, Art. no. e0130140, doi: 10.1371/journal.pone.0130140.

[26] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 618–626, doi: 10.1109/ICCV.2017.74.

[27] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001, doi: 10.1214/aos/1013203451.

[28] Q. Zhao and T. Hastie, "Causal interpretations of black-box models," *J. Bus. Econ. Statist.*, vol. 39, no. 1, pp. 272–281, Jan. 2021.

[29] A. Goldstein, A. Kapelner, J. Bleich, and E. Pitkin, "Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation," *J. Comput. Graph. Statist.*, vol. 24, no. 1, pp. 44–65, Jan. 2015, doi: 10.1080/10618600.2014.907095.

[30] J. Liu, N. Danait, S. Hu, and S. Sengupta, "A leave-one-feature-out wrapper method for feature selection in data classification," in *Proc. 6th Int. Conf. Biomed. Eng. Informat.*, Dec. 2013, pp. 656–660, doi: 10.1109/BMEI.2013.6747021.

[31] (2019). *LOFO Importance*. [Online]. Available: https://github.com/aerdem4/lofo-importance

[32] E. Štrumbelj and I. Kononenko, "Explaining prediction models and individual predictions with feature contributions," *Knowl. Inf. Syst.*, vol. 41, no. 3, pp. 647–665, Dec. 2014, doi: 10.1007/s10115-013-0679-x.

[33] A. Datta, S. Sen, and Y. Zick, "Algorithmic transparency via quantitative input influence: Theory and experiments with learning systems," in *Proc. IEEE Symp. Secur. Privacy (SP)*, May 2016, pp. 598–617, doi: 10.1109/SP.2016.42.

[34] M. Aslam, D. Segura-Velandia, and Y. M. Goh, "A conceptual model framework for XAI requirement elicitation of application domain system," *IEEE Access*, vol. 11, pp. 108080–108091, 2023, doi: 10.1109/ACCESS.2023.3315605.

[35] Z. Zhang, H. A. Hamadi, E. Damiani, C. Y. Yeun, and F. Taher, "Explainable artificial intelligence applications in cyber security: State-of-the-art in research," *IEEE Access*, vol. 10, pp. 93104–93139, 2022, doi: 10.1109/ACCESS.2022.3204051.

[36] A. Theissler, F. Spinnato, U. Schlegel, and R. Guidotti, "Explainable AI for time series classification: A review, taxonomy and research directions," *IEEE Access*, vol. 10, pp. 100700–100724, 2022, doi: 10.1109/ACCESS.2022.3207765.

[37] S. Bobek, M. Kuk, M. Szelazek, and G. J. Nalepa, "Enhancing cluster analysis with explainable AI and multidimensional cluster prototypes," *IEEE Access*, vol. 10, pp. 101556–101574, 2022, doi: 10.1109/ACCESS.2022.3208957.

[38] T. Nakanishi, "Approximate inverse model explanations (AIME): Unveiling local and global insights in machine learning models," *IEEE Access*, vol. 11, pp. 101020–101044, 2023, doi: 10.1109/ACCESS.2023.3314336.

[39] J. A. Hartigan, H. Spath, and J. V. Ryzin, "Clustering algorithms," *J. Marketing Res.*, vol. 18, no. 4, p. 487, Nov. 1981, doi: 10.2307/3151350.

[40] A. H. Ngu, J. Shen, and J. Shepherd, "Toward semantically meaningful feature spaces for efficient indexing in large image databases," *Manag. Multimedia Semantics*, pp. 1–29, 2005. [Online]. Available: https://www.igi-global.com/chapter/toward-semantically-meaningful-feature-spaces/25966

[41] S. Li, "Analysis of Sanda action based on CCNN model of IBS distance measurement," in *Proc. 3rd Int. Conf. Appl. Mach. Learn. Data Sci.*, vol. 1941, 2021, pp. 14–16.

[42] A. Entezami and H. Shariatmadar, "Damage localization under ambient excitations and non-stationary vibration signals by a new hybrid algorithm for feature extraction and multivariate distance correlation methods," *Struct. Health Monitor.*, vol. 18, no. 2, pp. 347–375, Mar. 2019, doi: 10.1177/1475921718754372.

[43] J. Opitz, P. Heinisch, P. Wiesenbach, P. Cimiano, and A. Frank, "Explainable unsupervised argument similarity rating with abstract meaning representation and conclusion generation," in *Proc. 8th Workshop Argument Mining*, 2021, pp. 24–35.

[44] H. S. A. Fang, N. C. Tan, W. Y. Tan, R. W. Oei, M. L. Lee, and W. Hsu, "Patient similarity analytics for explainable clinical risk prediction," *BMC Med. Informat. Decis. Making*, vol. 21, no. 1, p. 207, Dec. 2021, doi: 10.1186/s12911-021-01566-y.

[45] K. E. Papoutsakis and A. A. Argyros, "Unsupervised and explainable assessment of video similarity," in *Proc. Brit. Mach. Vis. Conf.*, 2019, pp. 1–17.

[46] Y.-S. Lin, Z.-Y. Liu, Y.-A. Chen, Y.-S. Wang, Y.-L. Chang, and W. H. Hsu, "XCos: An explainable cosine metric for face verification task," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 17, no. 3s, pp. 1–16, Oct. 2021, doi: 10.1145/3469288.

[47] A. Fisher, C. Rudin, and F. Dominici, "Model class reliance: Variable importance measures for any machine learning model class, from the," 1801, *arXiv:1801.01489*.

[48] G. König, C. Molnar, B. Bischl, and M. Grosse-Wentrup, "Relative feature importance," in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Jan. 2021, pp. 9318–9325, doi: 10.1109/ICPR48806.2021.9413090.

[49] S. Doyen, H. Taylor, P. Nicholas, L. Crawford, I. Young, and M. E. Sughrue, "Hollow-tree super: A directional and scalable approach for feature importance in boosted tree models," *PLoS ONE*, vol. 16, no. 10, Oct. 2021, Art. no. e0258658, doi: 10.1371/journal.pone.0258658.

[50] Y. Niu, C. Peng, and B. Liao, "Batch-wise permutation feature importance evaluation and problem-specific bigraph for learn-to-branch," *Electronics*, vol. 11, no. 14, p. 2253, Jul. 2022, doi: 10.3390/electronics11142253.

[51] K. Pearson, "On lines and planes of closest fit to systems of points in space," *London, Edinburgh, Dublin Phil. Mag. J. Sci.*, vol. 2, no. 11, pp. 559–572, Nov. 1901.

[52] M. Baisantry, A. K. Sao, and D. P. Shukla, "Discriminative spectral–spatial feature extraction-based band selection for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5518014, doi: 10.1109/TGRS.2021.3129841.

[53] Z. Li and Y. Qiu, "Feature selection based on improved principal component analysis," in *Proc. 2nd Asia Conf. Algorithms, Comput. Mach. Learn.*, New York, NY, USA, Mar. 2023, pp. 188–192, doi: 10.1145/3590003.3590036.

[54] M. Baisantry, A. K. Sao, and D. P. Shukla, "Two-level band selection framework for hyperspectral image classification," *J. Indian Soc. Remote Sens.*, vol. 49, no. 4, pp. 843–856, Apr. 2021, doi: 10.1007/s12524-020-01262-w.

[55] Y. Kiyoki, X. Chen, and H. Ohashi, "A semantic spectrum analyzer for realizing semantic learning in a semantic associative search space," *Front. Artif. Intell. Appl.*, vol. 136, pp. 50–67, Nov. 2006.

[56] R. E. Bellman, *Dynamic Programming*. Princeton, NJ, USA: Princeton Univ. Press, 1957.

[57] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 1–27, 2008.

[58] L. McInnes, J. Healy, and J. Melville, "UMAP: Uniform manifold approximation and projection for dimension reduction," 2018, *arXiv:1802.03426*.

[59] S. Lipovetsky, "PCA and SVD with nonnegative loadings," *Pattern Recognit.*, vol. 42, no. 1, pp. 68–76, Jan. 2009, doi: 10.1016/j.patcog.2008.06.025.

[60] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed. New York, NY, USA: Springer, 2002.

[61] Z. Wu, Y. Li, A. Plaza, J. Li, F. Xiao, and Z. Wei, "Parallel and distributed dimensionality reduction of hyperspectral data on cloud computing architectures," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 6, pp. 2270–2278, Jun. 2016, doi: 10.1109/JSTARS.2016.2542193. https://doi.org10.1109/jstars.2016.2542193.

[62] H. Zou, T. Hastie, and R. Tibshirani, "Sparse principal component analysis," *J. Comput. Graph. Statist.*, vol. 15, no. 2, pp. 265–286, Jun. 2006.

[63] N. Halko, P. G. Martinsson, and J. A. Tropp, "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions," *SIAM Rev.*, vol. 53, no. 2, pp. 217–288, Jan. 2011.

[64] J. Weng, Y. Zhang, and W.-S. Hwang, "Candid covariance-free incremental principal component analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 8, pp. 1034–1040, Aug. 2003.

[65] M. K. Vijaymeena and K. Kavitha, "A survey on similarity measures in text mining," *Mach. Learn. Appl., Int. J.*, vol. 3, no. 1, pp. 19–28, Mar. 2016.

[66] S. Aeberhard and M. Forina, "Wine," *UCI Mach. Learn. Repository*, 1991. [Online]. Available: https://archive.ics.uci.edu/dataset/109/wine

[67] W. Wolberg, O. Mangasarian, N. Street, and W. Street, "Breast Cancer Wisconsin (Diagnostic)," *UCI Mach. Learn. Repository*, 1995. [Online]. Available: https://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic

[68] (2008). *The 20 Newsgroups Data Set*. [Online]. Available: http://qwone.com/~jason/20Newsgroups/

[69] B. Lu, M. Charlton, C. Brunsdon, and P. Harris, "The Minkowski approach for choosing the distance metric in geographically weighted regression," *Int. J. Geographical Inf. Sci.*, vol. 30, no. 2, pp. 351–368, Feb. 2016, doi: 10.1080/13658816.2015.1087001.

TAKAFUMI NAKANISHI (Member, IEEE) was born in Ise, Mie, Japan, in 1978. He received the Ph.D. degree in engineering from the Graduate School of Systems and Information Engineering, University of Tsukuba, in April 2006. Since April 2006, he has been engaged in the research and development of knowledge cluster systems as well as text and data mining methods with the National Institute of Information and Communications Technology. In April 2014, he was appointed as an Associate Professor with the Global Communication Centre, International University. Since April 2018, he has been an Associate Professor with the Department of Mathematical Engineering, Faculty of Engineering. Since April 2019, he has been an Associate Professor with the Department of Data Science, Musashino University. His research interests include XAI, data mining, emotional information processing, and media content analysis.

• • •