# VisExPreS: A Visual Interactive Toolkit for User-driven Evaluations of Embeddings

Aindrila Ghosh, Mona Nashaat, James Miller, Shaikh Quader


**Fig. 1:** The Structural Quality Analysis View of VisExPreS. The interface is divided into ten regions that enable simultaneous and user-driven assessment of preserved structure in a group of low-dimensional embeddings. In the figure: region 'A' helps users to interactively select or upload datasets for analysis. Region 'B' enables the selection of desired dimensionality reduction algorithms. Region 'C' proactively guides users with the hyperparameter combinations for the chosen algorithms and assists them with further modifications of the hyperparameter values. Region 'D' provides proactive guidance on representative data-points and representative data subsets for analysis of embeddings. Region 'E' shows the embeddings as scatter plots where the users can interactively select data points for further analysis, visualize the neighborhoods of the selected points, and view the set of representative data-points that were proactively suggested by VisExPreS. The back-to-back bar graphs in Region 'F' show the feature influence explanations for the neighborhoods of the data-points under investigation. The buttons in region 'G' assist with exploration of different analytical aspects that VisExPreS has to offer. The bar graph and the donut charts in region 'H' allows users to interactively compute the local and global divergence metrics for the embeddings. Finally, region 'I' enables users to choose from options between the analysis of local structure, global structure, or features of the input dataset.

**Abstract**— Although popularly used in big-data analytics, dimensionality reduction is a complex, black-box technique whose outcome is difficult to interpret and evaluate. In recent years, a number of quantitative and visual methods have been proposed for analyzing low-dimensional embeddings. On the one hand, quantitative methods associate numeric identifiers to qualitative characteristics of these embeddings; and, on the other hand, visual techniques allow users to interactively explore these embeddings and make decisions. However, in the former case, users do not have control over the analysis, while in the latter case. assessment decisions are entirely dependent on the user's perception and expertise. In order to bridge the gap between the two, in this article, we present VisExPreS, a visual interactive toolkit that enables a user-driven assessment of low-dimensional embeddings. VisExPreS is based on three novel techniques namely PG-LAPS, PG-GAPS, and RepSubset, that generate interpretable explanations of the preserved local and global structures in embeddings. In the first two techniques, the VisExPreS system proactively guides users during every step of the analysis. We demonstrate the utility of VisExPreS in interpreting, analyzing, and evaluating embeddings from different dimensionality reduction algorithms using multiple case studies and an extensive user study.

**Index Terms**— Data and knowledge visualization, Interactive data exploration and discovery, Knowledge and data engineering tools and techniques.

## 1 Introduction

EMBEDDINGS are low-dimensional representations of high-dimensional data that are obtained using


- *Aindrila Ghosh, Mona Nashaat, and James Miller are with the Department of Electrical and Computer Engineering, University of Alberta, 116 St & 85 Ave Edmonton, Canada. E-mail: aindrila, nashaata, jimm@ ualberta.ca.*
- *Shaikh Quader is with IBM Toronto Software Lab, 8200 Warden Ave, Toronto, Canada, E-mail: shaikhq@ca.ibm.com.*


xxxx-xxxx/0x/$xx.00 © 200x IEEE Published by the IEEE Computer Society

Dimensionality Reduction (DR). DR algorithms transform high-dimensional data into embeddings while attempting to maximally preserve their structural properties [1]. DR as a transformation technique not only reduces the computational overhead of high-dimensional data analysis, but it also makes the visualization of such datasets possible with traditional spatial techniques (i.e., 2D or 3D plots).

Despite their utility, DR techniques come with a set of major caveats. Firstly, the dimensions derived using such techniques lack a clear-to-interpret mapping with the original features in the data [2], [3]. As a result, novice data analysts are often forced into blindly trusting the embeddings without truly understanding the projection axes or the positioning of data-points. Secondly, a plethora of DR techniques exist with their own respective hyperparameter combinations that significantly influence the embedding structure. The non-intuitive nature of these parameters also hinders the interpretability of these techniques making the selection of an appropriate DR method for any dataset, difficult [4]. Thirdly, in most cases, embeddings derived from DR do not make existing errors and distortions [5] prominent to users. In some cases [5]–[8], where such distortions are visually exposed, users are not allowed to control or interact with them [6], [7]. All these limitations make an efficient evaluation of embeddings obtained from DR algorithms extremely challenging [9].

Traditionally, the quality of embeddings is interpreted and evaluated using two different methods namely: (i) metric based quantitative analysis; and, (ii) qualitative or visual analysis of the obtained embedding. Being reliable and repeatable [9]–[11], metric-based quantitative evaluation of embeddings can effectively assist users to compare DR algorithms by associating numeric identifiers to their qualitative characteristics. **Nevertheless, such techniques being formally defined, do not allow users to have much control over the analysis process.** On the other hand, several qualitative analysis techniques for embeddings allow users to visually explore the neighborhood structures [12]–[14], errors & distortions [5]–[8], and feature variances [15], [16] within the neighborhoods of the projections. **Nevertheless, making any decision regarding the best performing DR algorithm using such techniques entirely depends on the analyst's perception and understanding of the embeddings.** Moreover, such techniques [2], [17], [18] often do not proactively guide users with the analysis process. For example, for very large datasets, most existing techniques [2], [17], [18] do not assist users with the selection of influential data points [13] or representative data subsets [19], [20] that provide a good representation of the original data and can reveal the overall quality of the embeddings better than other points. As a result, **novice data analysts often fail to utilize the entire potential of such interactive visualization techniques** when exploring low-dimensional embeddings.

**This research aims to bridge the gap between these two traditional techniques for evaluating embeddings.** In this work, unifying the benefits of both, we present a visual interactive toolkit that enables a proactively guided and user-driven analysis of preserved structure in any embeddings obtained from any DR algorithm. Towards achieving this goal, at first, we present two novel interactive embedding quality analysis methods. The first technique ***PG-LAPS (Proactively Guided Local Approximation of Preserved Structure)*** enables the computation of the *local-divergence*, that examines the fidelity of the relative positioning of any individual data-point in an embedding by approximating a neighborhood locally around that point. Moreover, to assist novice users with the analysis process, PG-LAPS proactively guides users with the selection of representative data points (i.e., analytically-interesting points such as outliers, center of clusters etc.) from the input dataset. Analysing such representative data points provides more valuable information [13] regarding the structural quality of an embedding than analysing any random point from the dataset. The second technique ***PG-GAPS (Proactively Guided Global Approximation of Projection Space)*** computes the *global-divergence*, that explains the preserved global structure in a low dimensional embedding, by combining non-redundant local-approximations from a coarse discretization of the projection space. To facilitate a proactively guided exploration of embeddings, as a part of PG-GAPS. we also present **RepSubset,** a novel algorithm that generates representative subsets from the original data based on the notions of density [13], [21] and dissimilarity [22], [23] in the dataset. These subsets aim to represent the structural properties of the original datasets as closely as possible so that they can be used with PG-GAPS for an accurate analysis of preserved global structure in embeddings. The three proposed techniques are then composed into a visual toolkit that we named ***VisExPreS (Visual Explanations of Preserved Structure)***. An overview of the VisExPreS interface is presented in Fig. 1. The presented toolkit not only gives users more control over the quality assessment process but also allows them to focus on the aspects of the analysis that are the most *interesting* to them. Moreover, VisExPreS enables side-by-side (both visual and quantitative) comparisons among multiple DR algorithms on the same dataset. Our primary contributions in this article are as follows:

1. *PG-LAPS*: a novel user-driven embedding quality analysis method for proactively guided investigation of the retained local structure in an embedding.
2. *PG-GAPS*: a novel, proactively guided, and user-driven DR quality assessment method for examining the preserved global structure in an embedding.
3. *RepSubset:* A novel algorithm for selecting representative subsets from the original dataset based on the notions of density and dissimilarity in the data.
4. *VisExPreS*: an interactive visual toolkit that enables a user-driven computation of metrics *local* and *global-divergence* metrics, while enabling side-by-side comparison of multiple embeddings.

The article is organized as follows: Section 2 presents the required background and related work; Section 3 discusses our design goals for VisExPreS interface. Section 4 presents PG-LAPS, PG-GAPS, and RepSubset and elaborates on the different views of VisExPreS. Section 5 presents an extensive evaluation of the methods; whilst Section 6 discusses some limitations and opportunities for future work. Section 7 concludes the article.

## 2 Background and Related Work

In its most general setting, non-linear DR (NLDR) can be formally defined as follows: assuming a matrix $X$ of size $n \times D$, that represents a high-dimensional dataset with $n$ records and $D$ features, DR algorithms map $X$ into an embedding $Y$ of size $n \times d$. For most real-world datasets $d$ represents the intrinsic dimensionality of $X$ that is the minimum number of dimensions that can be used to represent $X$ [5]. Normally, $d \ll D$. In their most general settings, NLDR techniques can be formally defined [5] as an optimization problem:

$$\underset{Y \in \mathbb{R}^{d \times n}}{\operatorname{argmin}} f(Y; X, \theta) \quad (1)$$

where the objective function $f$ attempts to preserve the relative proximities among the data points from $X$ to $Y$. In Eq. 1, $\theta$ represents the hyperparameters of the function $f$. In manifold learning [24], the vectors in $X$ are assumed to be sampled from a non-linear manifold where the notion of proximity among the data-points is traditionally defined using distance measures [1], [25]. In this research, for any data-point $x_i \in X$ where the neighborhood [26], [27] of $x_i$ is a subset $Z$ of $X$ containing $x_i$, we define the proximity between the point $x_i$ and its nearest neighbors as $\pi_{x_i}(x')$, where, $x' \in Z$ and $x_i \neq x'$. Here, $\pi_{x_i}(x') \in \mathbb{R}$. Ideally, for any NLDR, the preservation of the relative proximities among data-points refers to the following: considering a set of data-points $x_i, x_j, x_k \in X$ where, $i \neq j \neq k$, if $\pi_{x_i}(x_j) < \pi_{x_i}(x_k)$ then after the transformation $\pi_{y_i}(y_j) < \pi_{y_i}(y_k)$ should hold. However, NLDR being an optimization problem, its outcome often converges to a local-optima leading to the relative proximities not being retained for every data-point in $Y$. For most NDLR techniques, such relative proximities form the basis of both quantitative and visual evaluations of their resulting embeddings [9], [10]. In this section, at first, we elaborate on the existing mechanisms for quality assessments of embeddings. Next, we present the related work that focuses on obtaining representative subsets from datasets for further analysis.

### 2.1 Quality Assessments of Embeddings

In this section, we elaborate on related work that evaluates embeddings using quantitative or qualitative (visual) methods and discuss the novelty of VisExPreS, our proposed visual interactive toolkit.

#### *2.1.1 Quantitative Evaluation of Embeddings*

In theory, for most NLDR techniques, a quantitative analysis of the resulting embedding should be possible using two simple mechanisms. Firstly, by examining the value of the objective function $f$ upon convergence; and secondly, by performing an inverse transformation from $Y$ to $X$. Nevertheless, in real-life scenarios, neither of the above-mentioned techniques is often applicable. The reasons being: as the former technique can only be used for comparing different executions of the same algorithm [9], the later becomes infeasible for real-world datasets whose underlying manifold structures are unknown [4]. As a result, most existing techniques that attempt to quantify the characteristics of an embedding [28], either by examining the absolute (i.e., distance-based) or relative (i.e., rank-based) proximities among the data-points after the transformation.

Popular DR quality metrics [29] include Local Continuity Meta-Criterion (LCMC) [9], Trustworthiness and Continuity (T&C) [9], Mean Relative Rank Errors (MRRE) [29] among others. To evaluate the embedding quality, these metrics compare the ranks of the sorted distances in K-ary neighborhoods in $X$ and $Y$ [10]. Due to their similarities, Lee and Verleysen have unified these three metrics under a co-ranking matrix framework [9]. Also, the authors proposed several other rank-based DR quality metrics namely: mean K-ary neighborhood preservation ($Q_{nX}$ [9]), local quality criteria for K-ary neighborhoods ($Q_{local}$ [10]) and global quality criteria for the embedding ($Q_{global}$ [10]). The co-ranking framework primarily examines the average agreement between all K-ary neighborhoods in $X$ and $Y$ based on a matrix containing the ranks of pairwise distances [10] between the data-points. Apart from these, the metrics entropy and mutual information [30] and Spearman rank correlation [29], [30] are also popularly used to determine the preservation of topology in embeddings. On the other hand, Residual Variance [29] is a popularly used distance-based DR quality metric that computes a linear correlation between the absolute distances between any pairs of points in $X$ and $Y$. Other distance based metrics that focus on dissimilarities in neighborhoods [24] include neighborhood hits [24], projection precision score [6], and Spectral Overlap [30]. All these metrics examine the proportion of the neighborhood that is preserved in an embedding. Apart from the distance and rank based metrics, some other quality measures analyze the stress (i.e., distortion) of the DR objective function. Such metrics popularly include normalized stress [7] and Krushkal Stress [6], [29].

As shown by Lee et al. [10] and Johannemann et al. [30], such quantitative embedding quality measures can be useful for comparing the performances of multiple DR algorithms. However, these metrics being formally defined, do not allow much flexibility in the analysis process. Moreover, it can often get challenging for users to actively engage with their computations that will enhance the user's trust on the metric's value. For example, in most cases such metrics present a single quantitative value for the embedding quality, without providing any rationale behind the computation of the metric. In such cases, the users remain unable to intervene or interact with the metric computation process and are bound to trust the presented result blindly without really understanding how it was computed.

#### *2.1.2 Visual Evaluation of Embeddings*

In order to make the analysis of embeddings more engaging for users, several visual interactive mechanisms [2], [8], [17], [18], [31] have been proposed in the past few years. Among these, some techniques only visualize the neighborhood structures of the data-points after DR [12]. Some depict the distortions in the embeddings [7], [13], [14] with the help of false [13] and missing [7] neighbors. Taking the interaction one step further, some techniques even allow users to intervene and fix (i.e., reposition or remove) [2], [16] any misplaced data-points [5] in the embeddings. For example: as Smilkov et al. [12] effectively visualizes the

neighborhoods of any selected data-point in the embedding; Lai et al. [13], Cutura et al. [14], Martins et al. [7], and Aupetit et al. [31] identify and depict false and missing neighbors in embeddings. Moreover, using graphical representations, Heimerl et al. [8] show the distribution of neighborhood distances for all data points, whilst France et al. [32] depict the *agreements* in K-ary neighborhoods. On the other hand, in order to improve distortions, Stahnke et al. [5] and Joia et al. [33] automatically reposition misplaced datapoints; whereas, Pagliosa et al. [16] and Cavallo et al. [2] allow users to interact with the DR process and remove or relocate points in embeddings. Among further existing techniques, some assist users to interactively modify the hyperparameter combinations for DR algorithms [12]. Some other techniques [13], [15]–[18] help to visualize the contributions of the features in the relative positioning of data-points using feature weights [13], [18], feature correlations [17], or the variance in feature values [15], [16].

The case-studies [5], [12], [17] and user-studies [5] performed by the authors of the existing techniques show that such visual embedding quality analysis techniques can indeed help with an interactive exploration of the projection. However, as these techniques are not designed to quantify the behavioral characteristics of the embeddings, the decision on the best performing algorithm relies on the perception and expertise of the analyst. Moreover, most of these techniques focus on partial aspects of embedding quality. For example, either their focus lie with investigating false and missing neighbors [7], [14], or with the contributions of the original features in the dataset [15]–[18].

In this research, we attempt to unify the different focus areas of visual analysis of embeddings and address the challenges that we have discussed for both quantitative and qualitative assessment methods for DR.

### 2.2 Selecting Representative Subsets from Data

The existing methods for representative subset selection can be broadly categorized into two groups, namely ([19], [34]): clustering-based design and uniform design approaches. The clustering-based techniques aim at generating clusters from the original data followed by selecting a diversified subset from the generated clusters [20]. Such techniques can be further classified into [19], [34]: hierarchical, non-hierarchical, and density based techniques. Over the past years, for clustering-based subset selection popular methods such as K-means [20], OPTICS [33], DBSCAN [20] have commonly been used by researchers. For example, in their work, Lai et al. [13] have used DBSCAN to suggest representative data points as well as data subsets, whereas Daszykowski et al. [20] have suggested representative points using a hybrid of DBSCAN and K-Means clustering methods.

With the uniform design approaches [20]–[22], [35], the representative data-points are selected in such a way that they uniformly cover the data-space. Being more popular than the clustering-based subset selection methods, over the past few decades several approaches have been proposed for selecting uniform representative subsets. One of the most popular techniques in this category is the Kennard-Stone [23] algorithm that is based on the notion of *dissimilarity* between data-points. The Kennard-Stone is an iterative method that aims at minimizing the pairwise Euclidean distances [17] between the data-points that are already a part of the representative subset and the remaining points in the dataset. Another popular method, a dissimilarity based uniform subset selection technique is OptiSim, is presented by Clark et al. [22]. This technique generalizes the maximum and minimum dissimilarity-based approaches. Also based on the notion of dissimilarity, Tominaga et al. [35] presented a genetic algorithm to select representative subsets. The algorithm uses pairwise Euclidean distances among the data-points within the selected subset along with the mean of the product moment correlation as its two fitness functions. On the other hand, instead of focusing on dissimilarity among data points, researchers such as Chaudhuri et al. [21], [36] and Mall et al. [20] have focused on the idea of *density* of each point in the dataset for generating a representative subset from the data. For example, in a multidimensional space, Chaudhuri et al. [21] have presented a technique that selects data points with the highest density in the dataset. Whereas for connected graphs, Mall et al. [20] have presented the technique FURS that selects representative data-points with the highest measure of 'degree centrality' in the network. In both these iterative techniques, to enhance the diversity of the selected subset, the authors have ignored the K-nearest neighbors of the data points that have already been added to the representative subsets.

The representative subset selection technique RepSubset presented in this research, addresses multiple aspects that are not considered by the traditional iterative subset selection algorithms [20]–[22], [35]. First of all, to enable a diversified subset selection RepSubset combines the notions of density and dissimilarity in the original dataset while ignoring the impact of K-nearest neighbors of the already selected points. Secondly, in order to capture the nonlinearity in the underlying manifold of the original dataset, in contrast to the existing approaches, the proposed technique computes the pairwise geodesic distances [4] to measure the dissimilarities among data points.

## 3 Design Goals for the VisExPreS Toolkit

The primary goal for the VisExPreS toolkit is to assist users with an interactive and engaging quality assessment of low-dimensional embeddings. DR being a complex and black-box technique, this toolkit should not only guide users with the assessment process but also should allow users to have the driver's seat in the interactive and quantitative quality analysis of embeddings. In this section, we elaborate on our primary design goals for VisExPreS.

**Goal 1: Proactive guidance for representative data:** In order to assist novice data scientists with their analysis, the toolkit needs to provide proactive guidance to its users in terms representative data points [13] as well as, representative data subsets [22] for analyzing the preserved local and global structures of the embeddings. Among the two, the local structure represents the structural similarities in individual neighborhoods in the dataset, as the global structure signifies the relative differences in overall

neighborhoods in the original dataset [4]. In case of representative data-points, the toolkit should consider multiple perspectives from which a data-point may seem interesting and offer users several perspectives to select from. For representative subsets, the toolkit should automatically generate diverse representative subsets from the input data and present users with multiple subset options. Moreover, the toolkit should also suggest appropriate hyperparameter combinations for the DR methods.

**Goal 2: Simultaneous investigations of embeddings:** To address the challenges of selecting the most appropriate DR algorithm in a given scenario, we require the toolkit to be algorithm agnostic and assist with side-by-side comparisons among multiple embeddings.

**Goal 3: Contributions of the original features:** Features of a dataset play the most important role in computing the proximities among the points in a dataset. Moreover, as humans, we relate to meaningful names [37] more easily than numeric values or complex visuals. Hence, to enhance the interpretability of the analysis, we need the toolkit to interactively present the influences of the original features in the formation of the neighborhoods in the dataset.

**Goal 4: Interpretable assessment of structural quality:** We need the toolkit to address the non-transparent nature of the embeddings. Here, we require the toolkit to assist with interpretable explanations of preserved local and global structures in embeddings for both novice and expert data analysts. For this purpose, we need the toolkit to present multiple aspects of structural quality in embeddings annotated with textual descriptions.

**Goal 5: User-driven computation of quality criteria:** In this research, we define the term user-driven as follows: the computation of any quality metrics as a part of the interactive toolkit, should be completely user-steered. That is, the toolkit should put the users in charge of entire quality analysis process. It should not only let them select data points of their interests for the assessment, but it must also allow users to actively participate in determining the contributions (i.e., weights) of each component of the defined quality metrics. Based on this definition, to bridge the gaps between the quantitative and visual assessment methods for embeddings (cf. Section 2.1), we need the presented toolkit to be *user-driven* in nature.

## 4 The VisExPreS Framework

In par with the five design goals defined in Section 3, at the core of the *VisExPreS* (***Vis**ual **Ex**planations of **Pre**served **S**tructure*) toolkit we incorporate three novel techniques. These include two quality analysis techniques for embeddings and a representative subset selection method for large datasets. The first quality analysis technique *PG-LAPS* ***(P**roactively **G**uided **L**ocal **A**pproximation of **P**reserved **S**tructure)* allows for an interactive computation of the metric *local-divergence* that both visually and quantitatively assesses the local neighborhoods of individual data-points in embeddings. The second quality analysis technique *PG-GAPS* ***(P**roactively **G**uided **G**lobal **A**pproximation of **P**rojection **S**pace)* helps with the computation of a global quality criterion *global-divergence*. The metric global-divergence quantifies the preserved global structure for a set of non-redundant data points in the embedding. In VisExPreS, the set of non-redundant points is recommended by the toolkit using the incorporated RepSubset technique in the form of a representative subset of the original data.

It is important to note that, the techniques PG-LAPS and PG-GAPS are presented as enhancements of our previously proposed methods LAPS and GAPS [38]. In this research, we have added proactive guidance for selection of representative data-points (also data-subsets) and hyperparameters for the DR algorithms and have incorporated the two techniques into the visual interactive toolkit VisExPreS. Also, as apart of PG-GAPS we present RepSubset an novel method for selecting representative data subsets with high coverage. In the next two sections, at first, using Figures 2 and 3, we discuss the PG-LAPS and PG-GAPS processes along with the proposed RepSubset algorithm, in detail. More detailed discussions on the LAPS and GAPS methods can be found in [38]. Next, we justify the interface design for VisExPreS with respect to its design goals discussed in Section 3.

### 4.1 Proactively Guided Computation of Local-Divergence in Embeddings

As depicted in Figure 2, the Proactively Guided LAPS or PG-LAPS process can be divided into seven distinct steps. Overall, PG-LAPS investigates a single data-point in the dataset and enables users to quantify its preserved local structure using the output metric local-divergence. In the following, we discuss each step of PG-LAPS in detail.

**Step 1: Pre-process Input Data and Obtain Embedding:** As shown in Fig. 2- Step 1, to avoid additional noise in the obtained embeddings and also to enhance their stability [26], [27], [39], the pre-processing of our input data begins with an estimation of the intrinsic dimensionality $d$ (cf. Section 2). Here, following a popular practice in the literature [4], [26], [39], we use the maximum likelihood intrinsic dimensionality estimator for this purpose [39]. At the same time, we also perform an automatic and exhaustive grid search [4] to identify optimum hyperparameter combinations (cf. Eq. 1, parameter $\theta$) for the chosen DR method given the data. Next, the embedding for the input data is obtained by executing any DR algorithm chosen by the user using the estimated value for $d$ and the identified hyperparameter combinations.

**Step 2: Proactively Suggest Representative Data Points:** To enable the evaluation of preserved local structure in an embedding, PG-LAPS allows users to interactively select a single data point for subsequent analysis. Alternatively, as shown in Fig. 2 - Step 2, the proactive guidance component of PG-LAPS. suggests to users sets of representative data-points that might be interesting from five different analytical perspectives. Here, data-points are considered to be interesting if the point (1) is an outlier, (2) has a highly dense neighborhood, (3) is misplaced (i.e., has false or missing neighbors in the projection), (4) is close to the decision boundary, (5) is the center of any cluster. All these options

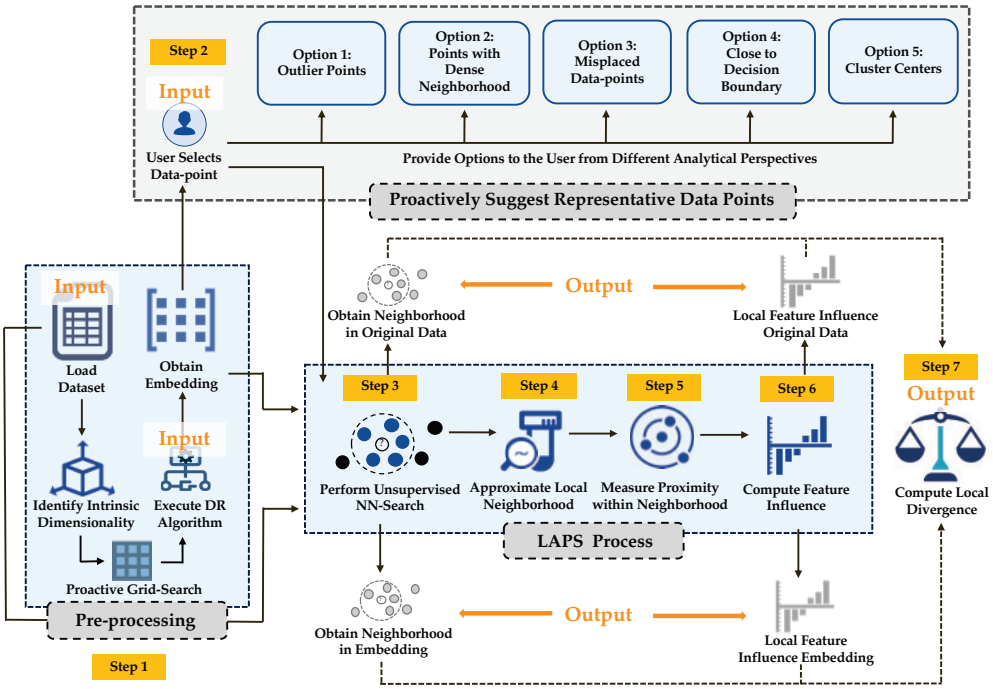


Fig. 2. Overview of the PG-LAPS technique. The technique primarily contains three components namely: Pre-processing, Proactive Guidance, and the LAPS process [38]. The overall process steps for PG-LAPS are numbered in an ascending order and are highlighted in yellow. Each of these process steps (i.e., Step 1 to Step 7 shown in the figure) are discussed in detail in Section 4.1.

are well-known aspects in academia for identifying influential data-points in any dataset [14].

Next, we discuss how the PG-LAPS system identifies representative data points from these five analytical perspectives, in detail.

- **Option 1:** The outliers in the original data are identified using a combination of the notions of density [13], [21] and dissimilarity [22], [23] among the data points. Here, at first, we measure the pairwise geodesic distances [5] among all data points. The geodesic distance $dist_\gamma$ among any two points $x_i$ and $x_j$ in $X$, can be defined as:

$$dist_\gamma(x_i, x_j) = infimum\{L(\sigma)\} \quad (2)$$

  where $inf\{L(\sigma)\}$ represents the infimum over the lengths of all the smooth paths $\sigma$ connecting the two points $x_i$ and $x_j$. Next, we arrange the points in a descending order of dissimilarity as we measure the neighborhood density of top 50% data-points with the maximum dissimilarity. The points with the lowest density are highlighted in the projections as potential outliers in the data.

- **Option 2:** Here, we use the DBSCAN ([13], [19] )algorithm to compute the density of all data-points and the data points with the highest density are presented to the user based on their chosen threshold. Density Based Spatial Clustering of Applications with Noise (DBSCAN) is a commonly used technique among researchers [13], [20] to identify data-points with dense neighborhoods.

- **Option 3:** For this option, we present users with the points with the lowest trustworthiness [9] in each embedding and allow them to further investigate these points. The metric trustworthiness [3], [9] measures the effect of false and missing neighbors on the rank [9] of data points in each neighborhood. Hence, in this option, we identify the points with the highest number of false and missing neighbors in the embedding.

- **Option 4:** Here, the data-points with the minimum distance from the decision boundaries are presented as representative points of interest to the users. It is important to note that, this option is only applicable for labelled datasets, where we use Support Vector Machine (SVM) for multi-class classification [40] in order to measure to the Euclidean distance of all data points from the decision boundaries.

- **Option 5:** In this option, the K-means clustering [19] technique is used to present only the cluster centers as representative data points to the users. Here, PG-LAPS presents the users with 'real' points from the dataset that are closest to the centroids of the data.

It is important to note that, since each of the different types of representative points are analytically interesting for different reasons, selecting points from different categories may produce different results for PG-LAPS. That is an DR algorithm that could preserve the outliers in it embedding better than others may not be the best in retaining the same cluster centers after the transformation.

As shown in Fig 2, once the user interactively selects an $x_i \in X$ (either among the proactively guided points or interactively from the dataset) for further analysis, the next steps of PG-LAPS are initiated.

**Step 3: Perform Unsupervised Nearest Neighbors Search:** In this step, as shown in Fig. 2, PG-LAPS simultaneously identifies the localities around the chosen point $x_i$ and its low-dimensional counterpart $y_i$ using an unsupervised $k$-nearest neighbor search. We formally define the identified local neighborhoods of size $k$ for $x_i$ and $y_i$ as $nn_{x_i}$ and $nn_{y_i}$ [38]. Next, the original feature vectors from $X$ are identified

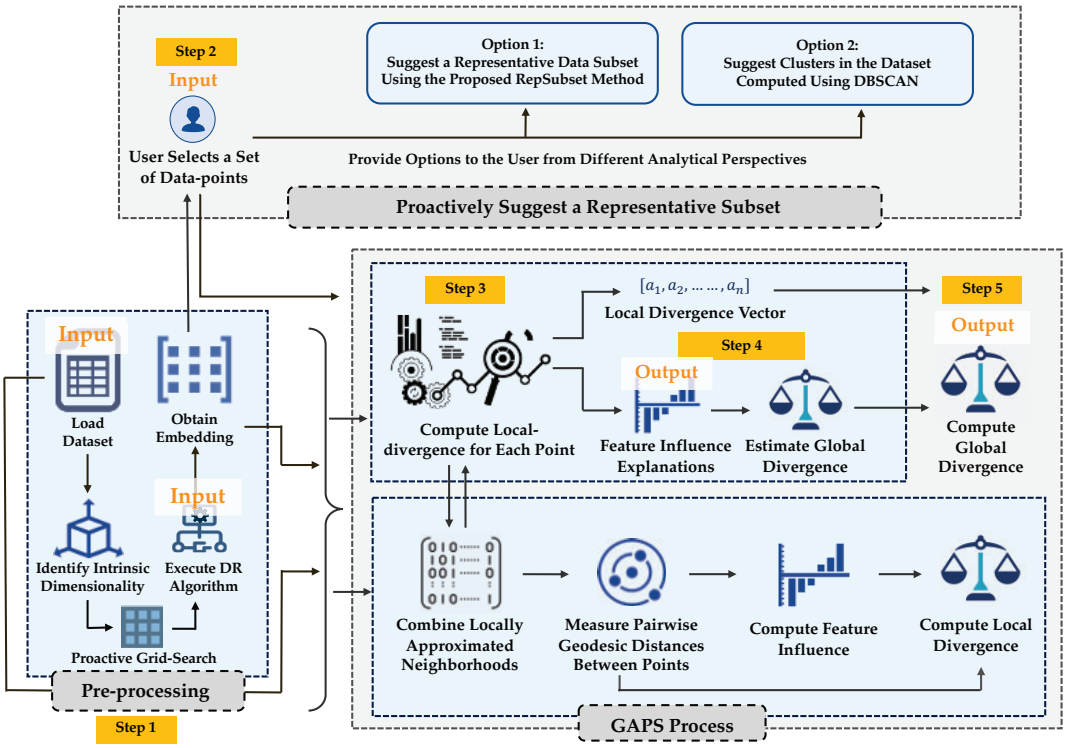


Fig. 3. Overview of the PG-GAPS technique. The technique primarily contains three components namely: Pre-processing, Proactive Guidance, and the GAPS process [38]. The overall process steps for PG-GAPS are numbered in an ascending order and are highlighted in yellow. Each of these process steps (i.e., Step 1 to Step 5 shown in the figure) are discussed in detail in Section 4.2.

from the indexes of the points in $nn_{x_i}$ and $nn_{y_i}$ and combined them into two feature vector matrices that we name $Z_{x_i}$ and $Z_{y_i}$ respectively [38].

**Step 4: Approximate Local Neighborhood:**
As the next step, the local neighborhood of the point $x_i$ and its low-dimensional counterpart $y_i$ are approximated by sampling a constant (user-defined) number of data-point samples with a normal distribution centered around each $x' \in Z_{x_i}$ and $y' \in Z_{y_i}$ [38]. The variances in the values of each attributes for the points in $nn_{x_i}$ and $nn_{y_i}$ were used as the sampling variance [38]. The primary reason for performing such an approximation is two-fold. Firstly, the process ensures local fidelity by amplifying the locality of the points $x_i$ and $y_i$ without having the need to consider an extremely large value for $k$. Secondly, such an approximation also ensures normality in the distribution of the feature values in the neighborhood. As a result, such perturbations of local neighborhoods is popular practice in literature and is used by authors such as Ribeiro et al. [37], Plumb et al. [41], and Guidotti et al. [42]. The approximated perturbed neighborhoods for $x_i$ and $y_i$ are combined into two feature vector matrices $\overline{Z_{x_i}}$ and $\overline{Z_{y_i}}$ respectively [38]. The size of $\overline{Z_{x_i}}$ and $\overline{Z_{y_i}}$ can be defined as $(k\ x\ M)\ x\ d$ considering $M$ points are sampled for each $k$ nearest neighbors of $x_i$ and $y_i$ .

**Step 5: Measure Proximities in the Local Neighborhood:**
Next, as shown in Fig. 2, the relative proximities $\pi_{x_i}(x')$ and $\pi_{y_i}(y')$ (cf. Section 2) are measured between the points $x_i$, $y_i$ and their perturbed neighbors respectively [38]. For feature vectors with continuous values, the Euclidean distance [17] is used as the proximity measure. Alternatively, in the case of feature vectors with a mixture of continuous and categorical values, Gower dissimilarity [5], [17], [43] is used to measure the proximity between them.

**Step 6: Compute Feature Influence Scores:**
In the following step (cf. Fig. 2), we compute the influences of the original features in the formation of the neighborhoods of $x_i$ and $y_i$. The primary reason behind this is that the influences of the original features in the dataset play a major role [7], [15], [16] in the structural preservation in embeddings. As, the relative proximities among the data points are identified by most DR algorithms from the original feature values in the dataset. Features can have either positive, negative, or no influence on the relative proximities among the data points. Whilst features with positive influences push the data-points further and hence, have a positive linear correlation with the increasing pairwise distances among data-points in the neighborhood [38]. Features with negative influences bring the data-points close to each other and have a negative (linear) correlation [38]. Similarly, non-influential features are those that show extremely low or no linear correlation with the increasing pairwise dissimilarities in the neighborhood. Detailed descriptions of the computation of feature influence explanations (cf. Fig. 2, step 6) $inf(x_i)$ for $x_i$ and $inf(y_i)$ for $y_i$ are presented in [38]. In the context of computing local divergence, $inf(x_i)$ and $inf(y_i)$ signify influences of the original features of the dataset on the relative dissimilarities between the data-points in the same neighborhood.

**Step 7: Compute Local Divergence**:
Finally, the metric local-divergence $\lambda_{x_i}$ for the selected data-point $x_i$ is computed as:

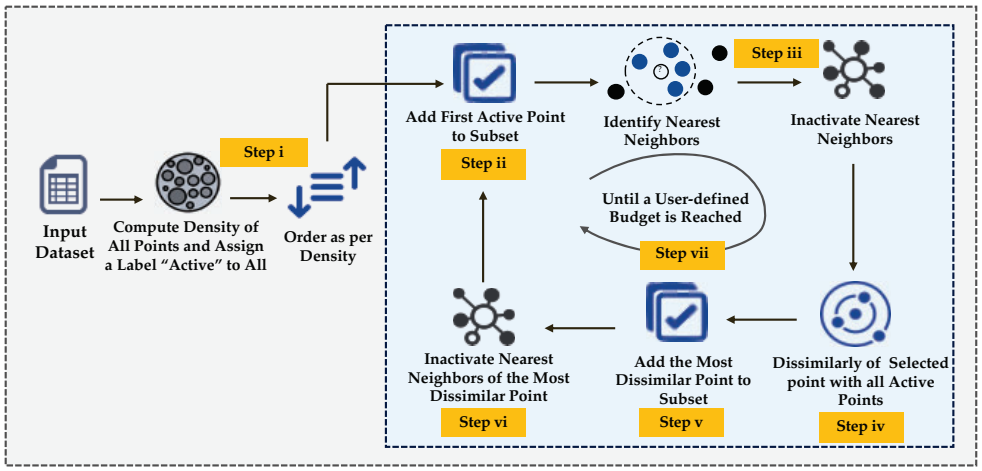


Fig. 4. Overview of the RepSubset technique. This novel iterative technique presents a representative subset of the input data as a part of the PG-GAPS process. The overall process steps for RepSubset are numbered in roman letters and are highlighted in yellow. Each of these process steps (i.e., Step i to Step ii shown in the figure) are discussed in detail in the Step 3 of Section 4.2.

$$\lambda_{x_i} = w_1 \pi_{\inf(x_i)}(\inf(y_i)) + w_2 \frac{nn_{x_i} \cap nn_{y_i}}{|nn_{x_i}|} + w_3 d_{r_{nn_{x_i}, nn_{y_i}}} \quad (3)$$

As shown in Eq. 3, local-divergence is composed as a weighted sum of three components. These include:

- $\boldsymbol{\pi_{inf(x_i)}}\boldsymbol{(inf(y_i))}$ **:** Signifies the *cosine* distance between $inf(x_i)$ and $inf(y_i)$ (cf. Step 6). That represents the discrepancy between the feature influences scores for the neighborhood of $x_i$ in the original dataset versus in the embedding.
- $\boldsymbol{nn_{x_i} \cap nn_{y_i}/|nn_{x_i}|}$ **:** Represents the false and missing neighbors for $x_i$ in the embeddings. Here we compute a ratio between the number of the preserved neighbors with the total number of $k$ nearest neighbors of $x_i$ considered for the analysis.
- $\boldsymbol{d_{r_{nn_{x_i}, nn_{y_i}}}}$ **:** Represents the difference between the relative orders (or ranks [9]) of neighbors in the neighborhoods of $x_i$ and $y_i$. This component measures whether an embedding can preserve the ordinal relationships among the data points in the original neighborhood.

In Eq. 3, $w_1$, $w_2$, and $w_3$ signify user-defined (scalar) weights of the three components of $\lambda_{x_i}$, by default in our computations, we consider each component of local divergence to be equally weighted (i.e., a weight of 0.33). However, during a user-driven computation of local-divergence the users are enabled to alter these weights in any way they seem fit. In either case $w_1$, $w_2$, and $w_3$ sum up to 1.

### 4.2 Proactively Guided Computation of Global-Divergence in Embeddings

The process of Proactively Guided GAPS (PG-GAPS) can be divided into five distinct steps as shown in Fig. 3. Overall, the process aims at examining the preserved global structure in an embedding. For this purpose, PG-GAPS suggests two possible types of representative subsets of the data and enables a user-driven computation of the metric global-divergence for any embedding. The detailed steps of PG-GAPS are as follows:

**Step 1: Pre-process Input Data and Obtain Embedding:** At the very beginning of the analysis, PG-GAPS performs similar pre-processing on the input dataset as discussed in Section 4.1. Here, as a part of the pre-processing, PG-GAPS estimates the intrinsic dimensionality $d$ of the input dataset followed by performing a proactive grid search of hyperparameters for the selected DR algorithms.

**Step 2: Proactively Suggest a Representative Subset:** The technique PG-GAPS analyzes the quality of the global structure in embeddings from a subset $X_S$ of non-redundant data-points in $X$. In order to perform a judicious selection of $X_S$, PG-GAPS presents users with proactive guidance for possible options. Although, users are allowed to interactively select data-points of their own choice around the manifold ignoring these guidelines. As shown in Fig. 3 PG-GAPS, as a part of the proactive guidance PG-GAPS recommends to users a representative subset of the original data that provides maximum coverage [20] of the input dataset. For this purpose, as a part of PG-GAPS, we present a novel representative subset selection technique RepSubset for diversified representative subset selection that is based on the notion of density and dissimilarities among the data points in the dataset. Here, we formally define the representative subset as $X_S \subseteq X$, such that $X_S$ provides a good representation of $X$ and $n_S \ll n$, where $n_S$ and $n$ are the sizes of $X_S$ and $X$ respectively. The step-by-step process of RepSubset is presented in Fig. 4 and is discussed below:

**Step i:** Compute the density of all data points in $X$. Order the points in terms of their decreasing density. Assign a status label of "active" to all points.

**Step ii:** Add the "active" point with the highest density to $X_S$. Update the status label of that point to be "inactive".

**Step iii:** Find the $k$-NN of the recently added data point. Update their status labels to be "inactive".

**Step iv:** Compute pairwise geodesic distances between the last added data point to $X_S$ and the remaining "active" points.

**Step v:** Select the point with the highest geodesic distance with the last added data point to $X_S$ and add the selected point to the $X_S$.

**Step vi:** Identify the $k$ nearest neighbors of the recently added data point to $X_S$ and update their status labels to be

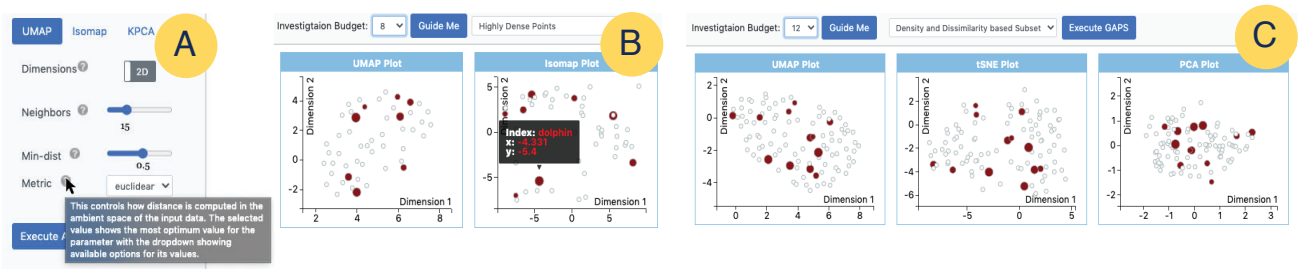


Fig. 5. Proactive guidance in the VisExPreS interface. The part A of the figure shows guidance on hyperparameter values in the embeddings. As the parts B and C of the figure show proactive guidance with representative data-points and representative data-subsets respectively.

"inactive".

**Step vii:** If $n_S < B$, go to step ii. Else, turn the status of all inactive points that do not already belong to $X_S$ to "active", and go to step ii.

As discussed in Section 2.2, the RepSubset technique is presented as an enhancement over the state-of-the-art subset selection methods. A detailed evaluation of the proposed algorithm is presented in Section 5.3. As an alternative to RepSubset, PG-GAPS also presents users with subsets of data-points that form clusters in the dataset. The purpose of this alternative is to provide users with multiple perspectives for selecting subsets from data. Here, following popular research [13], [20], the algorithm DBSCAN is used to identify clusters in the data. In both the cases, we let the users determine the number of instances they are willing to investigate and represent this number as an investigation budget $B$. Once the user selects their desired representative subset from the data, the next steps of the PG-GAPS process are initiated.

**Step 3: Compute Local-divergences of Points in Subset:** As shown in Fig. 3, after obtaining the data-points in $X_S$, we individually assess the local neighborhoods for each $x_i \in X_S$ and $y_i \in Y_S$ using the LAPS process [38] and compute their local divergence scores. This step is necessary because, in order to maintain an accurate global structure in the embeddings, the proximities among the points in the same neighborhoods should also be retained. These computed local-divergence scores for the data points in $X_S$ and $Y_S$ are then composed into two sets $\lambda_{X_S}$ and $\lambda_{Y_S}$ respectively [38]. Subsequently, we combine the perturbed local neighborhoods for each $x_i \in X_S$ and $y_i \in Y_S$ into two feature vector matrices $Z_{X_S}$ and $Z_{Y_S}$ [38].

**Step 4: Estimate the Global Divergence of the Subset:** A detailed description of estimating the global divergence of $X_S$ is presented in [38]. Here, we compute (pairwise) the pairwise geodesic distances among the points in the two feature vector matrices $Z_{X_S}$ and $Z_{Y_S}$. Next, the global feature distance contributions [44] and the feature influences are obtained in the same way as discussed in [38, p. 7]. Finally, we obtain an estimate of the global divergence $\lambda_{\bar{X_S}}$ for $X_S$ as a weighted sum of the disagreements in the overall estimation of the feature influences and the disagreements in the neighborhood structures of all points in $X_S$ and $Y_S$.

**Step 5: Compute Global Divergence Score:** Finally, a Global-Local Approximation (GLA) [45] is performed to obtain an additive blending of local approximations to form a globally-valid approximation (see [38, p. 7] for more details). To compute global-divergence, prior to the unification of the local-approximations, the ratios of the estimate of global divergence with the local-divergences for each points in $X_S$ are used as scaling factors [46] for each local-divergence score. As shown in Fig. 3, we define global-divergence as:

$$\lambda_X = \sum_{j=1}^{B} \frac{\lambda_{X_{S_j}}}{\lambda_{\bar{X}_S}} \lambda_{X_{S_j}} \quad (4)$$

where, $\lambda_X$ represents an additive blending of scaled local-divergence scores $\lambda_{x_i} \in \lambda_{X_S}$.

### 4.3 The VisExPreS Interface Design

As the final contribution of this research, we incorporated the two methods into a visual analysis toolkit, ***Vis**ual **Ex**planations of **Pre**served **S**tructure (**VisExPreS**)*. It is important to note that, although the VisExPreS toolkit can be used to compute other quality metrics for evaluating embeddings, the interface is primarily designed for PG-LAPS and PG-GAPS. The views in VisExPreS can be broadly categorized into two groups, namely: (1) Structural Quality Analysis View (cf. Fig. 1); and (2) Feature Analysis View. In this section, we justify the system design of the VisExPreS framework based on our design goals described in Section 3.

#### *4.3.1 Proactive Guidance for Representative Data*

As shown in Fig. 5, the interface of the VisExPreS toolkit is designed to provide proactive guidance in both the PG-LAPS and PG-GAPS processes as a part of its Structural Quality Analysis Views (cf. Fig. 1). This guidance is presented to the users in the following two ways. First of all, as depicted in Fig. 5a, the interface assists users with the selection of hyperparameter values for the chosen DR methods. Here, with a mouseover operation on the '?' icon next to the hyperparameter name, users are presented with a brief description of the hyperparameter itself. Alongside, using visual elements such as sliders and dropdown boxes (cf. Fig. 5), VisExPreS guides users with the possible values of the hyperparameter. Here, the optimum values of the hyperparameter for any chosen DR method and the corresponding dataset is presented as the default value of the hyperparameter by the interface. Whereas, the value ranges in the sliders and dropdown boxes represent other suitable values for the hyperparameter for the chosen dataset. At this point, the users can choose to proceed with the suggested value for the hyperparameters or can alter the values as per their choice.

Secondly, as shown in Figures 5b and 5c, VisExPreS

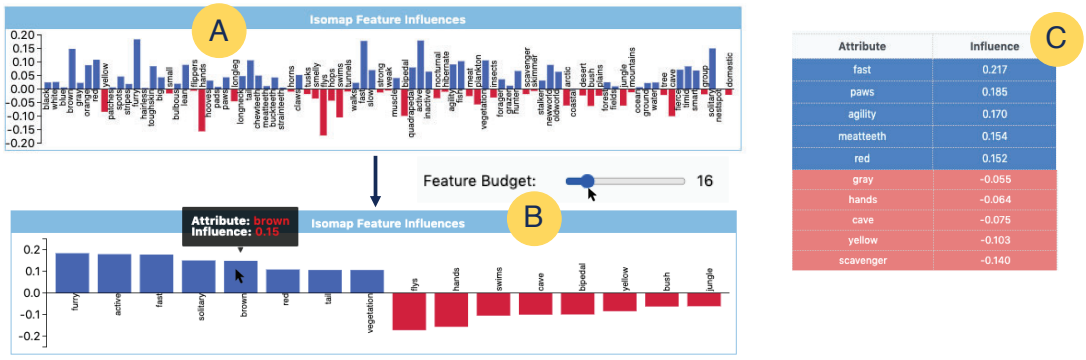


Fig. 6. Comparison of feature influences between the embeddings and the original dataset. The parts A and B show that by reducing the feature budget the users can investigate only most highly influential features in the dataset and compare them with the feature influences in the original dataset as shown in part C.

proactively guides users with the selection of representative data-points (i.e., for PG-LAPS) and data subsets (i.e., for PG-GAPS) when analyzing an embedding. Here, at first, users are enabled to select an investigation budget to limit the number of data-points that they want to see in the suggestions. This step is necessary as it can prevent the user from being overwhelmed with data-point options to select from. Once the investigation budget is selected, as depicted in Fig. 5b, for PG-LAPS, VisExPreS presents five different types (cf. Section. 4.1) of representative data-points to the users with the help of a drop-down list. Once the user makes a selection from one of these options, the respective representative data-points are visualized on the embeddings using the 'dark red' color and different radius sizes (cf. Fig. 5b and 5c). Here, the most representative data-point in the chosen category is presented with the largest radius size. The radius sizes of the other representative data-points gradually decrease based on their representativeness of their respective categories.

In the case of PG-GAPS, as shown in Fig. 5c, users are presented with two options for receiving suggestions on representative subsets (cf. Section 4.2). Similarly, as PG-LAPS, for PG-GAPS the data-points in the suggested data-subsets are highlighted with the color dark-red and the sizes of their radius show their representativeness in the subset. However, in the case of both PG-LAPS and PG-GAPS, users can choose to ignore the proactively suggested data-points, and interactively select their own point(s) of interest from the scatter plots of the embeddings for further analysis. In order to facilitate such an interactive selection of a data-point by the user, upon a mouseover operation, the interface presents users with contextual information for all points in the dataset. This include: the index, coordinates in the projection, class label, neighborhood contents, and neighborhood density of the points.

### 4.3.2 Simultaneous Investigations of Embeddings

In order to facilitate a side-by-side evaluation of embeddings, the VisExPreS interface presents the users with the option of making a selection of their choice of DR algorithms to compare on their choice of dataset (cf. Fig. 1, region-B). The design of the VisExPreS interface requires users to select at least three algorithms for this comparison. However, there is no maximum limit for the number of algorithms that can be compared. As shown in the regions C and E of Fig, 1, upon making a selection of the DR methods and deciding on the choice of their respective hyperparameter values, VisExPreS simultaneously presents the user with interactive scatter plots of the embeddings for the input dataset obtained from the chosen algorithms. Also, a mouseover operation on the embeddings show further details of individual data-points. Apart from the interactive scatter plots, all analysis results from PG-LAPS and PG-GAPS are presented simultaneously for the chosen DR methods. This allows users to perform a side-by-side comparison among the neighborhood structures of the data-points under investigation in multiple embeddings, as well as the influences of the original features (cf. Fig. 1, region-F) in the neighborhoods.

### 4.3.3 Contributions of the Original Features

As shown in Fig. 6, the contributions of the original features in the structural formation of the neighborhood of the chosen data-point is presented using back-to-back bar graphs for each chosen DR algorithm. The reason behind using back-to-back bar graphs for this purpose is to depict both positive and negative influences (cf. Section 4.1) of the features simultaneously. At the same time, the VisExPreS interface also shows the influences of the features in the original dataset (cf. Fig. 6c) allowing users to compare the differences between the original dataset and the embeddings. Upon looking at Fig. 6a and Fig. 6c, it can be argued that comparing the differences in feature influences can be difficult for datasets with a large number (i.e., <20) features. As shown in Fig. 6b, the VisExPreS interface provides a solution for this problem by allowing users to control the number of features that they are willing to investigate in the embeddings. Here, upon reducing the budget, the back-to-back bar graphs only show those features with the highest positive and negative influences in the structural formation of the neighborhood(s). At this point, the users can compare only those features with highest positive and negative influences (in the original data and in the embeddings) and check whether the embeddings have the same influences for the same features or not.

### 4.3.4 Interpretable Assessment of Structural Quality

The design of the VisExPreS interface aims to assist users to make the evaluation of the embeddings as interpretable as possible for both novice and expert users. For this

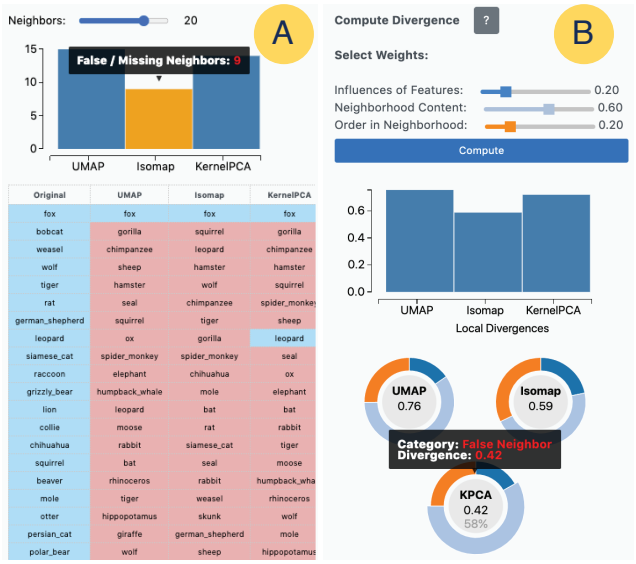


Fig. 7. (A) Neighborhood Analysis (B) Divergence computation in the VisExPreS Interface.

purpose, VisExPreS not only presents textual annotations for multiple aspects of analysis (cf. Fig. 6a), but also makes use of the visual interactive interface to assist users in a better engagement with the analysis process (cf. Fig. 6). For example, the colors in the back-to-back bar graphs are intentionally chosen to be bright enough so that they stand out of the text showing the attribute names (cf. Fig. 6a & 6b). Also, it is makes sure that in the back-to-back bar graphs the y-axes of each graph shows the same value range (cf. Fig. 6a & 6b). Thirdly, the neighborhoods of the selected points are enabled to be effectively compared using the tabular representation (cf. Fig. 7a) highlighting the order of the neighbors in the original dataset and in the embeddings. Here, with only a single glance at the table, users can notice the discrepancies in the neighborhood. For example, instead of looking at the indexes of the neighbors, users need to only check the colors in the table. As shown in Fig. 7a, the left-most column in the table shows the neighborhood of the chosen data-point in the original data and it is highlighted in sky blue. In this table, the user only needs to check in which other columns in the table there is another sky blue cell. Such a cell means that, the exact same neighbor and its respective order in the neighborhood was preserved in this embedding. Here, with just one glance at the table, the user can see which algorithm has preserved the neighborhood order (i.e., the third component of the local divergence metric) the best among all. As for the false and missing neighbors in the embeddings, VisExPreS presents a bar graph as shown in Fig. 7a. Finally, in case the differences are still not easily visible for the user, or if the users are not sure how the metrics local and global divergence are computed, the VisExPreS interface provides its users with two solutions. Firstly, as shown in Fig. 7b, upon a mouseover operation over the '?' icon next to the "Compute divergence" label, VisExPreS explains the rationale behind the computation of local or global divergences in detail. Secondly, as shown in Fig. 7b, VisExPreS presents the users with a set of three donut charts that quantify the discrepancies of the individual aspects (cf. Section 4.1) of the metrics local or global divergences.

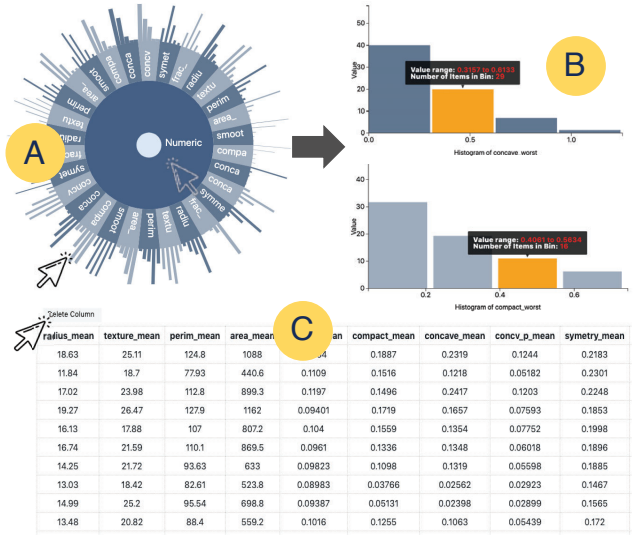


Fig. 8. The Feature Analysis View in the VisExPreS Toolkit. The part A shows a histogram amalgamated sunburst showing the categorization and value distributions of all numeric features in the dataset. Part B zooms into the histograms of individual features as part C shows all features and records in the dataset. The part C of the feature analysis view also allows to remove or add features into the analysis.

### 4.3.5 User-driven Computation of Quality Criteria

To support a user-driven assessment of embeddings, the VisExPreS interface puts the user in-charge of the analysis in several different ways. First of all, as shown in Fig. 1c, VisExPreS allows users to select the DR algorithms and their respective hyperparameter values of their choice. Secondly, as shown in Fig. 5b, the point(s)-of-interest to be chosen for the analysis of local and global structures completely depend on the user's preference. Thirdly, the computation of the metric for local and global divergence primarily depends on the user's decision of the relative weights for the three components of the metrics (cf. Section 4.1). As shown in Fig. 7b, with the help of the interactive sliders, users are allowed to modify the default weights of the different components of the metrics, this alters the values of the two-output metrics of PG-LAPS and PG-GAPS. Finally, the feature analysis view of the VisExPreS interface (cf. Fig. 8) allows users to visually analyze and interact with the features in the original dataset. Figure 8 shows an example of the feature analysis view in VisExPreS. The view can be divided into three primary regions as shown in this figure. Region A depicts a histogram amalgamated sunburst diagram that groups features based on their types (i.e., numeric or categorical) and shows the value distribution of each feature around the perimeter of the sunburst diagram. Here, the visual effectiveness of the sunburst is obtained by grouping the features into numeric and categorical ones. To utilize the visual effectiveness of the sunburst even more, we leave a more advanced grouping of the attributes [47] as our future work. Region B (in Fig. 8) zooms into the distributions of the individual features in the dataset and shows the histograms of their value distributions. Finally, region C shows the actual records in the

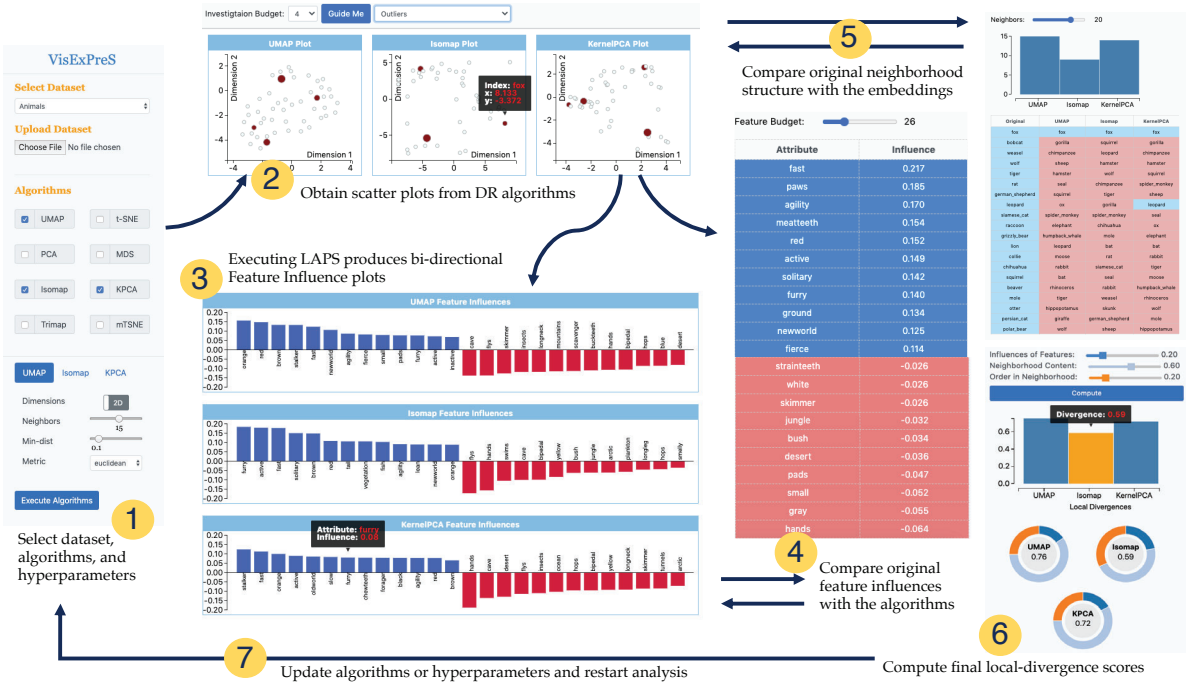


Fig. 9. Local structural quality analysis of the Animals Cancer dataset with PG-GAPS using VisExPreS. Each step of the assessment for the interactively chosen data-point 'fox' is highlighted with numbers and annotated with the details of the respective step.

data; the region also allows users to interactively remove one or more features from the analysis that may seem less influential to the user. Removal of such features would reduce the redundancy in the DR process that will improve the quality of the embeddings leading to lower local and global divergence scores for the algorithms.

## 5 Experimental Evaluation

In this section, we present the results of our detailed experimental evaluation of the VisExPreS toolkit as well as the proposed RepSubset algorithm. This section is primarily divided into two parts. In the first part, we perform an exhaustive evaluation of VisExPreS, whilst in the second part we compare RepSubset with some of its closest competitors. Further details regarding our experimental evaluation of VisExPreS are presented as supplemental materials.

### 5.1 Case Study of Data Analysis with VisExPreS

The VisExPreS interface is primarily designed to evaluate the quality of embeddings obtained by executing different DR methods on the same dataset. As a result, the target user type for the toolkit are data scientists who execute DR algorithms as a part of their day-to-day analysis of datasets and need to make decisions regarding which DR method would be more suitable for their subsequent analysis given their input dataset and analytical context.

In this section, we present an example of data analysis using the VisExPreS interface that demonstrates its utility in a user-driven evaluation of embeddings. For a better understanding of the toolkit's utility, another such example can be found in our supplemental material and a detailed demonstration video of embedding quality analysis with VisExPreS can be found at: https://bit.ly/3jax7Og

#### Analysis of the Animals Dataset using VisExPreS

Alice is a zoological data scientist who studies the behavioral patterns of wild animals and performs predictive modelling of their appearances in the wild. For this purpose, Alice has obtained an open source image dataset [17] that contains 30,475 images of 50 animals that are classified using 85 numeric features. As commonly performed by practitioners [4], [48], [49], Alice decides to perform DR on the input data prior to training a predictive model with it. Alice has limited experience with DR, but she is aware of that several algorithms exist for this purpose. At this point, Alice decides to explore the data using the VisExPreS interface (cf. Fig. 9). In order to enhance visual clarity of the embeddings, she uses a subset of 100 points from the dataset for her preliminary analysis.

Fig. 9 shows the flow of Alice's analysis on the Animals dataset. Once Alice loads her dataset into the VisExPreS system and selects three DR algorithms she has heard the most about (i.e., UMAP [48], Isomap [4], and KernelPCA [4]), VisExPreS executes a proactive grid search on her chosen algorithms for the input data and presents her with some suggestions for the hyperparameter values for the algorithms. Being a novice data scientist, Alice hovers her mouse pointer over the '?' icons next to these hyperparameters that explain the purpose of the parameter using tooltips (cf. Fig 9.1). The tooltips also state the optimum values for these hyperparameters for her dataset is already selected for her by the toolkit. Alice decides to proceed with the suggested values of these hyperparameters for her analysis.

Upon choosing to execute the DR algorithms, the VisExPreS interface presents Alice with three embeddings for her input dataset. Here, Alice decides to explore the proactive guidance from the framework in order to select a representative data-point from the dataset for further analysis (cf. Fig 9.2). With her investigation budget set at 4, Alice

looks for outliers in the dataset. VisExPreS identifies the points *dolphin*, *tiger*, *sheep*, and *fox* as outliers. However, Alice chooses to investigate the point *fox* for further analysis. Now, upon executing LAPS on point *fox*, the VisExPreS interface shows her with the neighborhood for the chosen point on the scatter plots as the feature influence explanation bar-graphs for the point are shown just below the scatter plots (cf. Fig. 9.3). Here, the VisExPreS interface allows Alice to compare the neighborhood and feature influences for *fox* in the embeddings and in the original dataset. Here, Alice notices that, the original feature influences presented in the right-side bar (cf. Fig. 9.4) show that, for *fox* the features *fast*, *paws*, *agility*, *meatteeth, red, active, furry,* and *solitary* have the highest positive influences. On the other hand, features *scavenger*, *yellow*, *cave* and *hands* have the most negative influences on its neighborhood structure. Some features such as *tunnel*, *bipedal*, and *forager* have little or no influences on the relative proximities in the neighborhood of *fox*. To investigate the back-to-back bar graphs, Alice reduces the feature budget in Fig. 9.4 to 16. Here, Alice only wants to investigate which of the embeddings have preserved similar influences for the most positive and most negatively influencing features in the original dataset for the neighborhood of *fox*. Upon carefully looking at the bar-graphs, Alice notices that Isomap has preserved the positive feature influences for features such as *fast, furry, red, agility,* and *solitary.* Also, in terms of the most negative feature influences Isomap has preserved the influences of *hands*, *yellow*, and *cave*. Whereas, in the embeddings produced by UMAP and KernelPCA, Alice cannot find many common features with high positive and negative influences as in the original data.

Upon comparing the neighborhood of *fox* in the original dataset (cf. Fig. 9.5) with the neighborhoods in the embeddings, Alice notices among the 20 neighbors visible in the scatter plots, Isomap has only 9 false and missing neighbors in its embedding, whereas both UMAP and KernelPCA has 14 false or missing neighbors. In terms of the preserved orders of neighbors, KernelPCA has preserved only the point *leopard* in its actual position in the original dataset. The two remaining algorithms have completely messed up the orders of the points in the neighborhoods.

At this point, Alice decides to compute the *local-divergence* score (cf. Fig. 9.6) for the chosen point. Hence, at first Alice looks into the local-divergence scores presented by the VisExPreS interface, where equal weights were allocated for all three components of the metric (cf. Section 4.1). Here, Alice hovers her mouse pointer on the '?' icon next to the "Compute divergence" label and learns about the calculation of the local-divergence for the chosen point. She also observes the donut charts, where she can see the contributions of each of these components in the final value of local divergence score for the three DR methods. Here, Alice notices that Isomap has the lowest local divergence score with the default weights. Upon increasing the weight for the neighborhood content component, Alice notices that Isomap still performs better than others. Even with a higher weight to the feature influence component, Isomap performs the best among the three. At this point Alice decides to analyze the point *dolphin* and compare UMAP, Isomap, PCA for this point. Alice repeats the analysis for 50 points in the dataset and notices that on an average Isomap has performed better than all other algorithms. So, she chooses to execute Isomap on her dataset prior to training her predictive model with the data.

TABLE 1
RESULTS OF USABILITY ANALYSIS ON VISEXPRES

| Analytical Aspects | Novice | Experts |
|---|---|---|
| Effectiveness - mean (SD) | 0.89 (0.06) | 0.96 (0.03) |
| Efficiency - mean (SD) | 55.92 (4.20) | 41.09 (8.32) |
| SUS Score - mean (SD) | 69.29 (6.07) | 83.00 (4.47) |
| Design Goals (Fleiss' Kappa) | 0.27 | 0.48 |

***Note:*** *The table above summarizes the results of our user study. Here we compute the usability metrics defined in ISO 9241-11 standard using 12 human subjects. Additionally, the last row presents the users' agreements on the fulfillment of the design goals of the VisExPreS interface.*

## 5.2 Usability Analysis of the VisExPreS Interface

In this section, we perform a detailed user study and analyze the usability of the VisExPreS interface. Based on the guidelines of Georgsson et al. [50], we used metrics in the International Organization for Standardization (ISO) 9241-11 standard[1] and quantify the usability using **effectiveness**, **efficiency**, and **satisfaction**. In this section, at first. we discuss the setup of our user study followed by quantitative analysis of its results. More detailed results of our analysis are presented as supplemental materials.

### 5.2.1 Experimental Setup and Participants

Following the guidelines of Stahnke et al. [5], Ribeiro et al. [37], and Georgsson et al. [50], we performed our user-study on VisExPreS using 12 human subjects. The study was performed under controlled conditions [50] where the study organizers observed and assessed the interactions of the study participants with the VisExPreS interface. The participants were carefully chosen as a group of graduate students and industry professionals with a strong background in computing. However, among the participants, only five members had some understanding of DR, as the remaining seven were completely unfamiliar with the topic. In this study, we considered the former five as expert users and the latter seven as non-experts or novice users. During this study, the participants were asked to analyze the Breast Cancer dataset [17] using the VisExPreS interface. Prior to the study, all participants were given a 60-minutes walk-through of the dataset along with the different functionalities of the VisExPreS framework. Next, the participants were given 75 minutes time to analyze at least 18 data-points from the proactively guided categories of PG-LAPS (cf. Section 4.1) and two rounds of global quality analysis using the representative subsets suggested by PG-GAPS. To be more specific, the participants were asked to investigate 4 points from the dataset belonging to each of the first four proactively guided categories of PG-LAPS (i.e., points that are outliers, have dense neighborhoods,

[1] https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en

TABLE 2
EVALUATION OF COVERAGE FOR REPSUBSET

| Datasets | Data Statistics | | Coverage (Cov) | | |
|---|---|---|---|---|---|
| | #Rows | #Feat. | RepSubset | FURS | OptiSim |
| Breast Cancer | 569 | 32 | 0.86 | 0.81 | 0.72 |
| Wine Quality | 4898 | 12 | 0.93 | 0.88 | 0.56 |
| Magic | 19020 | 11 | 0.83 | 0.84 | 0.67 |
| Credit Card | 30000 | 24 | 0.87 | 0.84 | 0.63 |
| Animals | 30475 | 85 | 0.91 | 0.89 | 0.76 |
| MNIST | 60000 | 784 | 0.64 | 0.51 | 0.37 |

have low trustworthiness, are close to the decision boundary) and 2 points for the last proactively guided category (i.e., cluster centers). Here, the participants were required to compare the preserved local structure for these points in any three embeddings and make a decision on the best performing DR method. For the two representative subsets, the participants were asked to perform global quality analysis of the same three embeddings for the Breast Cancer dataset and select the best quality embedding among the three. At the end of their analysis, the responses from the study participants were recorded and the usability metrics from the ISO 9241-11 standard were computed from them.

#### 5.2.2 Measuring Effectiveness of VisExPreS

According to the ISO 9241-11 standard, effectiveness is one of the most important attributes of usability that is measured using the metric *degree of task completion* [50]. Here we measured given the total number of tasks, how many of the tasks could be successfully completed by the participants. In our case, the total number of tasks that was given to each study participants were 20. Following the footsteps of Georgsson et al. [50], we encoded the task completion in three different ways as: (i) a score of 1, in case a task was completed by a participant without any assistance, (ii) a score of 0.5, for situations when a participant needed minor assistance from the study organizers to complete their task, (iii) a score of 0, when a participant could not complete a task. We summarize our analysis results in Table 1. Here, similarly as Georgsson et al. [50], we present the *mean* and *standard deviation (SD)* of task completion rate for novice and expert users. Considering, a task completion rate of 100% to be an ideal case. However, as a common consensus among researchers [50], any score above 78% signifies an acceptable rate of effectiveness in a system. As shown in Table 1, for the novice and expert users the mean task acceptance rate were 88% and 95% respectively. In both cases, the standard deviation was less than 0.06. This confirmed that the VisExPreS interface enabled a high level of efficiency for novice and expert users.

#### 5.2.3 Measuring Efficiency of VisExPreS

ISO 9241-11 standard has identified efficiency as the second important measure of a system's usability that is measured using the level of effort and the amount of resources that were used by study participants to complete as task. Again, following the guidelines Georgsson et al. [50], we measured efficiency using the amount of time (in minutes) that was spent by the study participants in order to complete each of the given tasks. For this purpose, we asked the participants to use a timer and record their overall duration of performing each individual task. It is important to note that, 15 minutes of analytical time were ignored from this computation as the response time for the VisExPreS interface. The second row in Table 1 shows the *mean* and *SD* of analysis time reported by the novice and expert users for all 20 tasks. Our results show that, overall the novice users took much longer than the expert users to finish the same analytical tasks. However, all users could complete the respective tasks within the given timeframe. A more detailed analysis of the efficiency of VisExPreS can be found in supplemental materials.

#### 5.2.4 Measuring User Satisfaction from VisExPreS

The third metric for usability that is identified by the ISO 9241-11 standard is user satisfaction. In our study, following the footsteps of Georgsson et al. [50], the System Usability Scale (SUS) developed and designed by Brooke [51]. SUS contains 10 pre-defined standard utility questions with a provision to answer each question on a 5-point Likert scale. In our study, the questions of SUS were scored based on the guidelines presented in its original definition [51]. That is, the score contribution of each item was designated from 0 to 4, where for the questions 1, 3, 5, 7, and 9, the score was considered to be the selected scale position minus 1. For questions 2, 4, 6, 8, and 10, the score allocated was five points minus the scale position. The sum of the scores of all 10 questions was then multiplied by 2.5 to get the overall satisfaction value ranging between 0 and 100. As per Georgsson et al. [50], a SUS score above 70 is considered as good whilst, a score of 85 or above is considered as an indicator of excellent usability. Table 1 shows the results of our analysis where we present the *mean* and *SD* of SUS scores for the novice and expert study participants. Our results show that although the novice users had some difficulty in using the VisExPreS interface with a SUS score close to 70, the expert users found the interface to be extremely useful for analyzing embeddings.

As an additional step of the usability analysis, we also quantified the users' agreement on the fulfillment of the pre-defined design goals for the VisExPreS interface that were discussed in Section 3. For this purpose, following the guidelines of Lewis et al. [52], we computed the Fleiss' Kappa ($\kappa$) consistency measure for the five pre-defined design goals. The value of $\kappa$ ranges from -1 to +1, where -1 represents no agreement, +1 signifies perfect agreement and 0 denotes agreement due to random chance. The results of our study are summarized in Table1. The table shows that although the novice users had low positive agreement (i.e., 0.27) on the complete fulfillment of the design goals due to their prior experience with DR, the expert users had a moderate agreement of 0.48 regarding the same. Detailed analysis of user agreements on the design goals is presented in our supplemental materials.

### 5.3 Evaluation of Coverage for RepSubset

In this section, we evaluate the presented RepSubset algorithm and compare its 'Coverage' [20] with two state-of-the-art representative subset selection methods. As mentioned by Mall et al. [20], coverage is a simple evaluation

TABLE 3
A DETAILED SCALABILITY ANALYSIS OF THE PG-LAPS AND PG-GAPS PROCESSES

| Datasets | Mean (SD) of execution time for 100 points using PG-LAPS | Mean (SD) of execution time for 100 executions using PG-GAPS | | | | |
|---|---|---|---|---|---|---|
| | | 5 points | 10 points | 50 points | 100 points | 500 points |
| Breast Cancer | 38.01 (1.95) | 66.80 (6.48) | 71.40 (5.32) | 129.30 (12.85) | 219.60 (18.15) | 217.00 (17.84) |
| Wine Quality | 12.83 (0.32) | 23.00 (3.15) | 24.10 (4.57) | 42.40 (6.24) | 81.30 (7.22) | 77.60 (8.44) |
| Magic | 11.42 (0.23) | 22.90 (4.85) | 24.60 (4.62) | 34.10 (5.74) | 38.40 (5.22) | 49.20 (6.21) |
| Credit Card | 32.62 (0.38) | 44.20 (4.23) | 50.90 (5.32) | 167.70 (25.85) | 294.20 (21.78) | 335.50 (27.03) |
| Animals | 65.54 (4.66) | 31.80 (4.19) | 34.30 (5.48) | 51.70 (7.88) | 128.50 (12.32) | 211.40 (18.62) |
| MNIST | 70.87 (9.89) | 24.10 (4.23) | 24.80 (4.99) | 37.50 (6.41) | 71.60 (9.47) | 133.70 (17.44) |

metric for subset selection algorithms which measures a ratio between the total number of unique points that can be reached against the total number of points. In this research, to measure the coverage of the RepSubset algorithm, at first, we set our required subset size to be 10% of the original dataset and count the number of unique data-points that are in the neighborhood (for a $k$ of size 15) of all the points in the subset. Then we compute the ratio of these unique data-points with the size of the entire dataset. We compare the coverage of RepSubset with FURS [20] and OptiSim [22], two well-known representative subset selection methods. The results of our comparisons along with the statistics for the 6 datasets that were used in our analysis are presented in Table 2. The table shows that RepSubset has consistently shown a higher coverage (ranging between 64% to 93%) than both FURS and OptiSim for all the 6 datasets.

## 6 DISCUSSION

In order to provide further clarity on the usability of the VisExPreS toolkit; in this section, at first, we analyze the scalability of the presented PG-LAPS and PG-GAPS methods. Next, we identify the limitations of the proposed toolkit and present some ideas for future work.

### 6.1 Scalability Analysis of PG-LAPS and PG-GAPS

In order to enable user-engagement, analytical speed is one of the primary requirements for any graphical user interface. When designing VisExPreS we kept this in mind. This section presents a detailed scalability analysis of the two techniques PG-LAPS and PG-GAPS that are at the core of the VisExPreS interface. Our analysis results are presented in Table 3. In this table, for PG-LAPS we present the *mean* and *SD* of the required end-to-end computation time for 100 data points from 6 different datasets. On the other hand, for PG-GAPS, we have also presented the mean and SD for 100 executions of global quality analysis using for sample sizes of 5, 10, 50, 100, and 500 of the same six datasets. Overall, our results show that understandably PG-GAPS has a much higher execution time than the PG-LAPS process for all the six datasets. However, for PG-GAPS, the execution time does not increase significantly for large changes in the sample size (i.e., especially for when the sample sizes increase from 100 to 500). The primary reason is, as mentioned in [38], in step 3 of the PG-GAPS process the size of the overall perturbed neighborhoods is kept fixed to 5000. Here, there are 100 samples in the subset, PG-GAPS randomly generates 50 perturbed neighbors for each point. Whereas for 500 points in the sample, PG-GAPS randomly generates only 10 perturbed neighbors for each point. As a result, we can see that on a personal computer with 8 GB RAM and a processor with 4 cores, the VisExPreS interface has shown response time of a minimum 49.2 seconds and a maximum of 5.5 minutes for 500 data-points in the selected subset. However, we think that this range of the response time for PG-GAPS also depends on the number of features in the input dataset. Due to the finite scope of this work, we envision generating a best-practice guideline on the dataset volume and the number of features in support for a better scalability of the PG-LAPS and PG-GAPS processes as our future work.

### 6.2 Limitations and Future Work

VisExPreS being primarily based on spatial visualizations (i.e., 2D or 3D plots), all visual scalability limitations associated with special techniques also become applicable to the framework. For example, with more than 50 features in the input datasets, the bi-directional feature influences bar graphs or the histogram amalgamated sunbursts for feature analysis can become challenging to comprehend. Although this limitation cannot be completely avoided, the current implementation of VisExPreS allows users to reduce the feature budget to only investigate the most influential features. Also, the feature analysis view enables users to delete uninfluential or redundant features. However, we are currently working on enhancing the visual effectiveness of the sunburst in the feature analysis view by performing a more advanced grouping among the features in the dataset and allowing users to remove or alter groups of features as a whole. Finally, although VisExPreS successfully enables an interactive user-driven assessment of embeddings, it does not provide much assistance with driving the DR algorithms to enhance the embedding quality. Hence, as our ongoing work we are investigating on optimizing the local and global divergence metrics in order to improve the quality of the obtained embeddings.

## 7 CONCLUSIONS

This paper presents VisExPreS, a visual interactive toolkit that assists with a user-driven quality analysis of preserved local and global structures in the embeddings. At the core of VisExPreS, there are two novel techniques that are also introduced in this article. The first technique PG-LAPS generates interpretable explanations of the preserved locality of a single data-point in an embedding. PG-LAPS obtains the explanations regarding the structural preservation by

approximating the local neighborhood around the single data point that is proactively recommended by the framework. On the other hand, the second technique PG-GAPS explains the preserved global structure in an embedding by unifying the local approximations for a set of non-redundant data-points interactively selected from the projection space into a global approximation. To provide proactive guidance for the non-redundant data points required in the analysis as a part of PG-GAPS, we present RepSubset. A novel algorithm that uses the notions of density and dissimilarity among data-points to generate a representative subset from the data. We demonstrate the utility and usability of the proposed VisExPreS toolkit, PG-LAPS and PG-GAPS techniques, along with the RepSubset algorithm using an exhaustive evaluation. A large amount of our experimental results is presented as supplemental material.

## REFERENCES


[1] M. Sedlmair, M. Brehmer, S. Ingram, and T. Munzner, "Dimensionality Reduction in the Wild: Gaps and Guidance," Dept. Comput. Sci., Univ. British Columbia, Vancouver, Canada, Tech. Rep. TR-2012-03, Jun. 2012.

[2] M. Cavallo and Ç. Demiralp, "A Visual Interaction Framework for Dimensionality Reduction Based Data Exploration," *arXiv:1811.12199 [cs]*, Nov. 2018, Available: http://arxiv.org/abs/1811.12199.

[3] L. G. Nonato and M. Aupetit, "Multidimensional Projection for Visual Analytics: Linking Techniques with Distortions, Tasks, and Layout Enrichment," *IEEE Trans. Visual. Comput. Graphics*, vol. 25, no. 8, pp. 2650–2673, Aug. 2019, doi: 10.1109/TVCG.2018.2846735.

[4] L. van der Maaten, E. O. Postma, and H. J. van den Herik, "Dimensionality Reduction : A Comparative Review," *J Mach Learn Res 10*, vol. 66, no. 71, p. 13, 2008.

[5] J. Stahnke, M. Dork, B. Muller, and A. Thom, "Probing Projections: Interaction Techniques for Interpreting Arrangements and Errors of Dimensionality Reductions," *IEEE Trans. Visual. Comput. Graphics*, vol. 22, no. 1, pp. 629–638, Jan. 2016, doi: 10.1109/TVCG.2015.2467717.

[6] T. Schreck, T. von Landesberger, and S. Bremm, "Techniques for Precision-Based Visual Analysis of Projected Data," *Information Visualization*, vol. 9, no. 3, pp. 181–193, Sep. 2010, doi: 10.1057/ivs.2010.2.

[7] R. M. Martins, D. B. Coimbra, R. Minghim, and A. C. Telea, "Visual analysis of dimensionality reduction quality for parameterized projections," *Computers & Graphics*, vol. 41, pp. 26–42, Jun. 2014

[8] F. Heimerl, C. Kralj, T. Möller, and M. Gleicher, "embComp: Visual Interactive Comparison of Vector Embeddings," *arXiv:1911.01542 [cs]*, Nov. 2019, Available: http://arxiv.org/abs/1911.01542.

[9] J. A. Lee and M. Verleysen, "Quality assessment of dimensionality reduction: Rank-based criteria," *Neurocomputing*, vol. 72, no. 7–9, pp. 1431–1443, Mar. 2009.

[10] J. A. Lee and M. Verleysen, "Scale-independent quality criteria for dimensionality reduction," *Pattern Recognition Letters*, vol. 31, no. 14, pp. 2248–2257, Oct. 2010.

[11] A. Bibal and B. Frenay, "Measuring Quality And Interpretability Of Dimensionality Reduction Visualizations," In Safe ML Workshop at ICLR 2019.

[12] D. Smilkov, N. Thorat, C. Nicholson, E. Reif, F. B. Viégas, and M. Wattenberg, "Embedding Projector: Interactive Visualization and Interpretation of Embeddings," *arXiv:1611.05469 [cs, stat]*, Nov. 2016, Available: http://arxiv.org/abs/1611.05469.

[13] C. Lai, Y. Zhao, and X. Yuan, "Exploring high-dimensional data through locally enhanced projections," *Journal of Visual Languages & Computing*, vol. 48, pp. 144–156, Oct. 2018, doi: 10.1016/j.jvlc.2018.08.006.

[14] R. Cutura, S. Holzer, M. Aupetit, and M. Sedlmair, "VisCoDeR: A Tool for Visually Comparing Dimensionality Reduction Algorithms," *Comput. Intelligence*, p. 6, 2018.

[15] R. R. O. D. Silva, P. E. Rauber, R. M. Martins, R. Minghim, and A. C. Telea, "Attribute-based Visual Explanation of Multidimensional Projections," *EuroVis Workshop on Visual Analytics (EuroVA)*, p. 5 pages, 2015.

[16] L. Pagliosa, P. Pagliosa, and L. G. Nonato, "Understanding Attribute Variability in Multidimensional Projections," in *2016 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, Sao Paulo, Brazil, Oct. 2016, pp. 297–304, doi: 10.1109/SIBGRAPI.2016.048.

[17] M. Dowling, J. Wenskovitch, J. T. Fry, S. Leman, L. House, and C. North, "SIRIUS: Dual, Symmetric, Interactive Dimension Reductions," *IEEE Trans. Visual. Comput. Graphics*, vol. 25, no. 1, pp. 172–182, Jan. 2019.

[18] J. Z. Self, L. House, S. Leman, and C. North, "Andromeda: Observation-Level and Parametric Interaction for Exploratory Data Analysis.", Technical report, Department of Computer Sc., Virginia Tech, Virginia, 2015.

[19] M. Daszykowski, B. Walczak, and D. L. Massart, "Representative subset selection," *Analytica Chimica Acta*, vol. 468, no. 1, pp. 91–103, Sep. 2002.

[20] R. Mall, R. Langone, and J. A. K. Suykens, "FURS: Fast and Unique Representative Subset selection retaining large-scale community structure," *Soc. Netw. Anal. Min.*, vol. 3, no. 4, pp. 1075–1095, Dec. 2013.

[21] B. B. Chaudhuri, "How to choose a representative subset from a set data in multi-dimensional space," *Pattern Recognition Letters*, vol. 15, no. 9, pp. 893–899, 1994.

[22] R. D. Clark, "OptiSim: An Extended Dissimilarity Selection Method for Finding Diverse Representative Subsets ‡," *J. Chem. Inf. Comput. Sci.*, vol. 37, no. 6, pp. 1181–1188, Nov. 1997, doi: 10.1021/ci970282v.

[23] R. W. Kennard and L. A. Stone, "Computer Aided Design of Experiments," *Technometrics*, vol. 11, no. 1, pp. 137–148, Feb. 1969.

[24] R. Motta, R. Minghim, A. de Andrade Lopes, and M. C. F. Oliveira, "Graph-based measures to assist user assessment of multidimensional projections," *Neurocomputing*, vol. 150, pp. 583–598, Feb. 2015.

[25] S. Lespinats and M. Aupetit, "CheckViz: Sanity Check and Topological Clues for Linear and Non-Linear Mappings," *Computer Graphics Forum*, vol. 30, no. 1, pp. 113–125, Mar. 2011, doi: 10.1111/j.1467-8659.2010.01835.x.

[26] J. Xia, F. Ye, W. Chen, Y. Wang, W. Chen, Y. Ma, and A.K. Tung, "LDSScanner: Exploratory Analysis of Low-Dimensional Structures in High-Dimensional Datasets," *IEEE Trans. Visual. Comput. Graphics*, vol. 24, no. 1, pp. 236–245, Jan. 2018.

[27] K. Bunte, M. Biehl, and B. Hammer, "Dimensionality reduction mappings," in *2011 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)*, Paris, France, Apr. 2011, pp. 349–356.

[28] M. Espadoto, R. M. Martins, A. Kerren, N. S. T. Hirata, and A. C. Telea, "Towards a Quantitative Survey of Dimension Reduction Techniques," *IEEE Trans. Visual.*

*Comput. Graphics*, pp. 1–1, 2019.

[29] B. Rieck and H. Leitte, "Agreement Analysis of Quality Measures for Dimensionality Reduction," in *Topological Methods in Data Analysis and Visualization IV*, H. Carr, C. Garth, and T. Weinkauf, Eds. Cham: Springer International Publishing, 2017, pp. 103–117.

[30] J. Johannemann and R. Tibshirani, "Spectral Overlap and a Comparison of Parameter-Free, Dimensionality Reduction Quality Metrics," *arXiv:1907.01974 [cs, stat]*, Jul. 2019, Available: http://arxiv.org/abs/1907.01974.

[31] M. Aupetit, "Visualizing distortions and recovering topology in continuous projection techniques," *Neurocomputing*, vol. 70, no. 7–9, pp. 1304–1330, Mar. 2007, doi: 10.1016/j.neucom.2006.11.018.

[32] S. L. France and U. Akkucuk, "A Review, Framework and R toolkit for Exploring, Evaluating, and Comparing Visualizations," *arXiv:1902.08571 [cs, stat]*, Feb. 2019, Available: http://arxiv.org/abs/1902.08571.

[33] P. Joia, F. V. Paulovich, D. Coimbra, J. A. Cuminato, and L. G. Nonato, "Local Affine Multidimensional Projection," *IEEE Trans. Visual. Comput. Graphics*, vol. 17, no. 12, pp. 2563–2571, Dec. 2011.

[34] W. Gani and M. Limam, "A kernel distance-based representative subset selection method," *Journal of Statistical Computation and Simulation*, vol. 86, no. 1, pp. 135–148, Jan. 2016, doi: 10.1080/00949655.2014.996758.

[35] Y. Tominaga, "Representative subset selection using genetic algorithms," *Chemometrics and Intelligent Laboratory Systems*, vol. 43, no. 1–2, pp. 157–163, Sep. 1998.

[36] D. Chaudhuri, C. A. Murthy, and B. B. Chaudhuri, "Finding a subset of representative points in a data set," *IEEE Trans. Syst., Man, Cybern.*, vol. 24, no. 9, pp. 1416–1424, Sep. 1994, doi: 10.1109/21.310520.

[37] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why Should I Trust You?': Explaining the Predictions of Any Classifier," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, Feb. 2016, pp. 1135–1144, Accessed: Jul. 12, 2019. [Online]. Available: http://arxiv.org/abs/1602.04938.

[38] A. Ghosh, M. Nashaat, J. Miller, and S. Quader, "Interpretation of Structural Preservation in Low-dimensional Embeddings," *IEEE Trans. Knowl. Data Eng.*, pp. 1–1, 2020, doi: 10.1109/TKDE.2020.3005878.

[39] E. Levina and P. J. Bickel, "Maximum Likelihood Estimation of Intrinsic Dimension," *Advances in neural information processing systems*, pp. 777–784, 2005.

[40] C. Angulo, X. Parra, and A. Català, "K-SVCR. A support vector machine for multi-class classification," *Neurocomputing*, vol. 55, no. 1–2, pp. 57–77, Sep. 2003.

[41] G. Plumb, D. Molitor, and A. Talwalkar, "Model Agnostic Supervised Local Explanations," *arXiv:1807.02910 [cs, stat]*, Jul. 2018, Accessed: Aug. 12, 2019. [Online]. Available: http://arxiv.org/abs/1807.02910.

[42] R. Guidotti, A. Monreale, S. Ruggieri, D. Pedreschi, F. Turini, and F. Giannotti, "Local Rule-Based Explanations of Black Box Decision Systems," *arXiv:1805.10820 [cs]*, May 2018, Accessed: Aug. 12, 2019. [Online]. Available: http://arxiv.org/abs/1805.10820.

[43] J. C. Gower, "A General Coefficient of Similarity and Some of Its Properties," *Biometrics*, vol. 27, no. 4, pp. 857–871, 1971, doi: 10.2307/2528823.

[44] H. Wang and M. Hong, "Distance Variance Score: An Efficient Feature Selection Method in Text Classification," *Mathematical Problems in Engineering*, vol. 2015, pp. 1–10, 2015, doi: 10.1155/2015/695720.

[45] R. Haftka, "Combining global and local approximations," *AIAA journal*, vol. 29, no. 9, p. 1523, Sep. 1991.

[46] J. P. Boyd, "Additive blending of local approximations into a globally-valid approximation with application to the dilogarithm," *Applied Mathematics Letters*, vol. 14, no. 4, pp. 477–481, 2001.

[47] S. Mahmood and K. Mueller, "Taxonomizer: Interactive Construction of Fully Labeled Hierarchical Groupings from Attributes of Multivariate Data," *IEEE Trans. Visual. Comput. Graphics*, vol. 26, no. 9, pp. 2875–2890, 2020.

[48] E. Becht, L. McInnes, J. Healy, C.A. Dutertre, I.W. Kwok, L.G. Ng, F. Ginhoux, and E.W. Newell, "Dimensionality reduction for visualizing single-cell data using UMAP," *Nature Biotechnology*, vol. 37, no. 1, pp. 38–44, Dec. 2018, doi: 10.1038/nbt.4314.

[49] E. Amid and M. K. Warmuth, "A more globally accurate dimensionality reduction method using triplets," *arXiv:1803.00854*, 2018, www.arxiv.org/abs/1803.00854

[50] M. Georgsson and N. Staggers, "Quantifying usability: an evaluation of a diabetes mHealth system on effectiveness, efficiency, and satisfaction metrics with associated user characteristics," *J Am Med Inform Assoc*, vol. 23, no. 1, pp. 5–11, Jan. 2016, doi: 10.1093/jamia/ocv099.

[51] J. Brooke, "SUS - A quick and dirty usability scale," Usability evaluation in industry. Jun 11 1996, pp. 189.

[52] J. M. Lewis and V. R. de Sa, "A Behavioral Investigation of Dimensionality Reduction," *In Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 34, no. 34, p. 7, 2012.

**Aindrila Ghosh** is a Ph.D. student at the University of Alberta. She has obtained her Master's degree in Computer Science from the University of Paderborn. As her graduate research, she is working on enhancing the interpretability of machine learning models, along with improving user-engagement in the machine learning process.

**Mona Nashaat** received her BS and Master's degrees in Computer Engineering from Faculty of Engineering, Port Said University. She is now working toward the Ph.D. degree in the Department of Electrical and Computer Engineering, University of Alberta. Her research focuses on Machine Learning, Intelligent Systems, and Big Data.

**Dr. James Miller** P.Eng(Alberta) – jimm@ualberta.ca – has been a full professor with the Dept. Electrical and Computer Engineering at The University of Alberta since 2000. Previously, he was a professor at the University of Strathclyde (U.K.) and a principal research scientist at the National Electronics Research Initiative (U.K.). He has been an active researcher for more than thirty years across a wide range of topics, ranging from Computer Vision, Pattern Recognition, Embedded System Design, Software Engineering, Web Engineering and Proactive Analytics. He has published more than 100 articles in peer-reviewed journals including many IEEE and ACM venues.

**Shaikh Quader** received his B.Sc. in Computer Science from the University of New Brunswick followed by a Master's degree in Computer Science from the University of Waterloo. Shaikh is a Lead AI architect in IBM Canada. He leads research collaboration between IBM and academia and has published at several IEEE conferences.