International Society of the Learning Sciences

Challenges and Design Opportunities in Data Analysis for ML-Empowered Scientific Inquiry - Insights from a Teacher Professional Development Study

Xiaofei Zhou, Jingwan Tang, Beilei Guo, Hanjia Lyu, Zhen Bai
xzhou50@ur.rochester.edu, jtang21@u.rochester.edu, bguo5@u.rochester.edu, hlyu5@ur.rochester.edu, zbai7@ur.rochester.edu
University of Rochester

Abstract: The increasing impact of machine learning (ML) on real-world scientific investigations demands a shift to teaching ML-empowered scientific inquiry to familiarize younger students with authentic science practices. Evidence from previous studies has suggested that glyph-based data visualization is promising for making multidimensional data analysis accessible for novice learners. Little research, however, has examined the contextualized challenges of applying glyphs in K-12 ML-empowered scientific inquiry. Since teachers play a critical role at the frontier of technology integration for learning, as an initial attempt to address this research gap, we investigated the pattern recognition and interpretation behaviors of 18 K-12 STEM teachers when engaged in ML-empowered scientific inquiry during a professional development workshop. Findings confirm that teachers identified and interpreted ML-revealed patterns for scientific hypothesis formulation while it is challenging for novices to analyze and synthesize different patterns conceptually and extensively for more sophisticated inference. Underlying design opportunities and implications are discussed.

Introduction

In real-world science, an increasing number of scientists are applying machine learning (ML) to discover patterns from multidimensional data (Kitano, 2016). The majority of the current K-12 STEM teaching practices, however, focus on bivariable experiments and data analysis, which may limit students' learning opportunities of data analysis to investigate complex scientific phenomena (Kuhn, 2016). Recognizing and interpreting relevant patterns such as category, relationship, trend, and anomaly discovered by ML methods on multidimensional datasets may offer unique learning opportunities for evidence-based science practices aligned with the NGSS science and engineering practices (States, 2013). It is challenging, however, for K-12 students and teachers to recognize and interpret ML discovered patterns in science contexts due to their lack of ML and data literacy. Glyph-based data visualization (Ward et al., 2015), which maps multidimensional attributes of an individual data point to vivid visual features of everyday objects such as faces, stars, and trees, offers promising opportunities to make ML-discovered patterns accessible to novice learners. Glyph-based visualization supports visual reasoning (e.g., search, pattern recognition) over multidimensional datasets with high accuracy (Chung et al., 2015; Keck et al., 2017). There is a neglect of existing studies of glyph-based data visualization on supporting K-12 students and teachers for efficient data interpretation, an important skill for gaining insights into what the visuals show (Rau, 2017).

Teachers play a critical role in integrating technology for learning (Koehler & Mishra, 2009), which is especially important for integrating ML in STEM education. Therefore, as a first attempt to tackle the above research gaps, we (1) propose cluster analysis as an accessible ML technique for K-12 scientific inquiry and (2) examine how K-12 STEM teachers identify and interpret patterns from datasets using a glyph-based visualization tool named SmileyDiscovery (Zhou et al., 2021) in a professional development workshop. SmileyDiscovery utilizes face-based glyphs for data analysis with *k*-means clustering while requiring few efforts in learning ML. We collected log data of what patterns of scientific phenomenon (e.g., ecosystem, heart disease) are identified by 18 in-service K-12 STEM teachers using SmileyDiscovery, and analyzed how teachers interpreted these patterns to form scientific hypotheses. Teachers also provided feedback on potential learning experiences from the student perspective. Our research questions include **RQ1**: To what extent can teachers identify and interpret patterns during cluster analysis for scientific inquiry with face-based glyphs? **RQ2**: What challenges prevent teachers from effective scientific inquiry? The main contributions are:

- A framework on how cluster analysis can be an accessible discovery tool for K-12 teachers and students to conduct exploratory analysis on multidimensional data for scientific inquiry.
- Challenges in ML-empowered scientific inquiry by using face-based glyph visualization.
- Design implications to empower novice learners to analyze ML-generated multidimensional results.

Background

Multidimensional data analysis and visual representation in K-12 Education

Researchers have discovered positive results in developing middle school students' multivariable reasoning skills (Kuhn et al., 2015), training primary and high school students to make informal statistical inferences by modeling with multidimensional datasets (Braham & Ben-Zvi, 2017; Kazak et al., 2021), and facilitating high school students to identify multidimensional patterns from ML-generated results (Wan et al., 2020). These works show young learners' potential to apply ML as an advanced multidimensional data analysis tool for scientific inquiry and highlight the need to enrich more authentic scientific inquiry skills for K-12 STEM learning (Nayak et al., 2016) .

Visual representation facilitates multidimensional data analysis but can confuse students if they do not fully understand the visualized mechanisms and concepts (Rau, 2017). Previous studies have shown that full competence in visual representation comprehension is not achieved even by college and university graduates (Nayak et al., 2016; Roth et al., 1999). To develop students' representational competencies, Rau (2017) argued that two types of knowledge and skills were necessary for students to acquire: (a) visual understanding and fluency, referring to knowledge of general principles about how the visual representation depicts concepts and abilities to efficiently identify the corresponding conceptual meanings reflected by visual features; (b) connectional understanding and fluency, indicating an understanding of conventions for using multiple visual representations together and abilities to communicate conceptual implications of connections (i.e., similarities and differences) between different visual features.

Glyph-based visualization

Glyphs are well-suited for depicting high-dimensional data due to their capabilities of simultaneously displaying information through multiple visual channels, including shape, size, and color (Chung et al., 2015). Several studies have created and shown the effectiveness of glyph-based visualization in promoting multidimensional data analysis. Cao et al. (2011) presented an interactive system, DICON, which combined treemap-liked icons with Voronoi icons to illustrate multivariate data clusters. A case study demonstrated it supported cluster exploration and interpretation of a large dataset in the healthcare domain. Chung et al. (2015) designed a novel glyph and validated it with the analysis of Rugby matches. Findings confirmed its effectiveness in supporting multidimensional visual search and comparative analyses of multiple attributes between different matches. Keck et al. (2017) developed a flower glyph to facilitate big data analyses via clustering algorithms. An experimental study showed that it was more efficient than the traditional tabular displays for task completion and the accuracy for both conditions was almost the same. However, these positive effects are all about pattern recognition through data visualization, while little research has investigated whether or not the glyph approach enhances data analysis in relation to the mapped concepts. Therefore, this study attempts to gain a comprehensive understanding of how glyph design facilitates multidimensional pattern recognition and interpretation. It analyzed data from a study engaging k-12 teachers in pattern exploration enhanced by Chernoff faces (Chernoff, 1973) via *k*-means clustering. Chernoff faces aim to make multidimensional data accessible for human beings since people can easily recognize faces and notice changes in facial features (e.g., nose length, eye radius, and eyebrow slant, *etc*). Previous studies examined the visual effectiveness of Chernoff faces on displaying patterns (e.g., Lee et al., 2003; Morris et al., 2000), but did not address its interpretability in the K-12 STEM contexts.

A theoretical framework of applying cluster analysis for scientific inquiry in K-12 STEM education

Cognitive benefits of applying ML for scientific inquiry

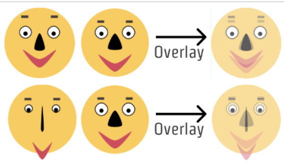
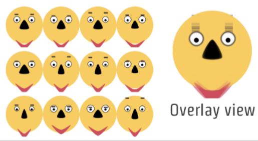
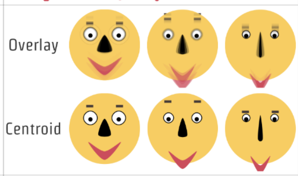
Pattern recognition could support inductive reasoning for scientific inquiry, which is a bottom-up probabilistic reasoning process that derives general principles of target phenomena from a large amount of observation (Rodrigues, 2011). Traditionally, induction plays a key role in hypothesis testing through accumulating evidence (Folger & Stein, 2017) for scientific inquiry. With ML revealing patterns from big data, learners' capability of inductive reasoning can be significantly enhanced not only during the investigation but also at the early stage of generating inquiry or formulating scientific hypotheses (Muller et al., 2016). In addition, ML analysis is usually accompanied by reasoning with contrastive explanation (Folger & Stein, 2017; Miller, 2019), which refers to the generation of explanation when encountering contrastive cases sharing similarities as well as differences in properties (Chin-Parker & Bradner, 2017). Contrastive explanation generates key insights from the puzzling event that occurred during the scientific inquiry (Folger & Stein, 2017; Miller, 2019).

Cluster analysis for inquiry-based learning in K-12 STEM education

As one of the most common ML techniques for data mining, cluster analysis has broad affordances to serve as an initial investigation for diverse directions to further analyze multidimensional data. By grouping similar data objects, it presents big data in a more accessible way, reveals inter-/intra-cluster patterns for further investigation, and discovers early insights for characterization, classification, outlier analysis, *etc* (Han et al., 2011). Thus, cluster analysis has been widely applied by researchers to explore datasets, bringing unique benefits in creating a scientific question, and/or formulating a scientific hypothesis (Romesburg, 2004), which connects with the conceptualization phase in inquiry-based learning for STEM education (Pedaste et al., 2015). For example, after thousands of multidimensional data points about ecological factors collected from field sites are grouped into different clusters, researchers can form hypotheses by contrasting the most dissimilar centroids or make inferences inductively on data clusters, and then investigate with another dataset. Besides the cognitive benefits of inductive reasoning and contrastive explanation, the main mathematical concepts behind clustering (i.e., centroid and distance-based clustering) are relatively intuitive for young learners to readily grasp (Greenwald et al., 2021). Furthermore, the skill to interpret trends and patterns from data could be aligned with NGSS (Next Generation Science Standards). Therefore, we attempted to propose clustering as an accessible and versatile inquiry-based learning tool for K-12 STEM education.

Figure 1

Cluster analysis for conceptualization in inquiry-based learning (Anderberg, 2014; Romesburg, 2004)

1. Pattern revealed by clustering	2. Glyph representation of the pattern in SmileyDiscovery	3. How does the pattern can support conceptualization in inquiry-based learning? Creating a scientific question	Formulating a scientific hypothesis
Similarities and dissimilarities between two data objects		E.g., cluster multivariate dental formulas of different animals. The result shows that the mole and the pig have identical formulas. What research hypothesis could explain this?	E.g., cluster multivariate dental formulas of different animals. The result shows that the walrus' formula is very different from the other carnivores. Does Walruses feeding habit accordingly modified their dental structure?
Intra-cluster pattern: Similarities and variations		E.g., cluster different species of primates based on 34 skin attributes. Intra-cluster pattern shows that the grouping is related to the features of the dermis, epidermis, and sweat glands, which is intriguing for further study.	E.g., within a relatively homogeneous group of ecological field sites created by clustering, sites with lower temperature have fewer small mammals. For this type of ecosystems, is higher temperature less livable for small mammals?
Inter-cluster pattern		E.g., cluster multivariate dental formulas and the sizes of different animals. The inter-cluster differentiating features show no apparent relation between an animal's dental formula and its size. Why?	E.g., cluster different wines based on their chemical factors and quality scores. The inter-cluster patterns show that wines with better quality have lower volatile acidity and more alcohol, which could be a hypothesis for further testing.
Clustering results with different sets of features	The overlay of centroids from two rounds of clustering 	E.g., cluster different animals based on (1) the diets of animals, (2) shapes of the teeth, (3) dental formulas. The clusters from the first clustering are more similar to the clusters from the second round than the clusters from the third. Why?	E.g., cluster different people by their income levels with two sets of other factors separately. The first round of clustering generates a better inter-cluster seperation in income levels. Does this set of factors (education, working hours) have a stronger influence on income?

Inspired by previous studies on cluster analysis (Anderberg, 2014; Han et al., 2011; Romesburg, 2004), we summarized a list of main patterns revealed by clustering which can be utilized for conceptualization in inquiry-based learning (see Fig. 1 column 1 & 3). Cluster analysis can help to formulate hypotheses for questions whose truth entails (dis)similarities among data objects. For example, cluster labels assigned to data objects could support the hypothesis of what data attributes lead to such grouping. Intra- and inter-cluster patterns could generate insights to hypothesize on correlations among data attributes for factor analysis or analysis of variance. Clustering with different sets of data attributes can suggest a set of data attributes distinguishing across the groups to a sufficient degree to aid discriminant analysis and classification. Besides assisting to form hypotheses, these patterns revealed by clustering could trigger further scientific questions. Considering the pedagogical and cognitive benefit of cluster analysis in inquiry-based learning, we utilized face-based glyphs to make cluster analysis an engaging and accessible tool for K-12 STEM teachers and students (Fig. 1 column 2). And this work is a preliminary evaluation of the initial design.

Methods and analysis

Research participants and procedure

We recruited 18 in-service K-12 STEM teachers who were taking a course named Integrating Technology with STEM Teaching, at a research-based university. They have very limited CS/ML backgrounds, different subjects, and various teaching grades (Table 1). The study session was conducted via an online meeting platform Zoom.

Table 1
Subjects and grade levels teachers teach

	Elementary school	Middle school	High school
Science	N = 2	N = 3	N = 6
Math	N = 2	N = 2	N = 3

In this study session, teachers participated in one cycle of scientific inquiry through cluster analysis on a dataset of field sites described by a set of ecological factors (Zhou et al., 2021), involving five steps: (1) generate the concept mapping for the face-based glyph (i.e., mapping each ecological factor to a facial feature) and get familiar with the value mapping (i.e., how the value of an ecological factor changes the corresponding facial feature); (2) compute the similarities between two representative ecological field sites; (3) explore the dataset by manually clustering a few data points around two representatives and interpret intra- and inter-cluster patterns; (4) apply the well-studied *k*-means clustering and interpret intra- and inter-cluster patterns; (5) formulate a hypothesis on how ecological factors dynamically influence each other based on the cluster analysis results. More detailed demonstration of this ML-empowered scientific discovery learning activity can be found in our previous work introducing the SmileyDiscovery system (Zhou et al., 2021).

During the study session, each teacher worked individually on their own device by following the instructional text in SmileyDiscovery, without any external guidance from researchers. Four teachers didn't complete the activity due to technical issues and their data were excluded. In the end, we collected teachers' feedback on potential learning experiences from students' perspectives.

Data analysis

RQ1*: To what extent can teachers conduct cluster analysis for scientific inquiry with face-based glyphs?* We employed system log data (i.e., clicking behavior and text input) and teachers' feedback to answer this question. First, we replicated teachers' clicking behaviors to revivify the concept mapping (i.e., how data attributes are mapped with facial features) and their learning experience. Second, we open coded (Strauss & Corbin, 1990) the text input and feedback on pattern recognition to measure *patterns recognized or neglected by teachers.* Third, we coded teachers' text inputs of scientific hypotheses drawn from clustering-revealed patterns to measure the *quality (i.e., accuracy and sufficiency) of interpretations of these patterns for forming scientific hypotheses.*

RQ2*: What challenges prevent teachers from conducting effective scientific inquiry?* Teachers described difficulties that they encountered when engaging scientific inquiry with face-based glyphs in both text inputs and feedback. We conducted a thematic analysis of the data and generated challenges that prevent effective pattern recognition and interpretation, respectively.

Three researchers coded data independently, met regularly to address disagreements, and refined codes.

Results

RQ1: To what extent can teachers conduct cluster analysis for scientific inquiry with face-based glyphs?

Novice learners' pattern recognition via face-based glyphs

By looking into the system log data recording teachers' clicking behaviors and text input of the patterns they identified for each step in the learning activity, we analyzed how novice learners identified patterns from face-based glyph visualization. During the step where teachers explored the dataset by manually clustering a few data points based on two representatives, some teachers engaged with more fail-and-trial experience by exploring the dataset, selecting more data points to test out the similarities, and removing the dissimilar ones. Other teachers carefully scanned through the dataset and only selected a few faces that were highly similar to the target face. While analyzing the patterns identified and reported by different teachers, we noticed inconsistent tolerance for variations, both across teachers and across facial features. For example, some teachers have a relatively high tolerance for variations and thus report more data attributes as shared patterns within clusters; teachers with a low tolerance for variations tended to identify most attributes as differentiating features. In addition, we found wrong/missing translations from facial features to data attributes, indicating the high cost of code-switching between science context and glyph visualization.

Five advantages of applying face-based glyphs to identify ML-revealed patterns are mentioned by teachers. First, teachers pointed out the pedagogical benefit of using emojis to visualize multidimensional data, which can be both familiar and playful for young learners: "This could be a fun way to allow students to explore data sets and draw connections in a way that allows for visualization". Second, teachers found that being able to customize the data visualization helps them to have a deeper understanding of different data attributes and how they form a glyph: "I tried to mentally pick ones that matched, like when I see a beetle my eyes get big like I don't want to see one, so to me was eye radius. Or latitude, I used the eyebrow height." Third, teachers appreciated the computation efficiency of using one piece of the glyph to represent a multidimensional data point and overlaying two pieces to compute the (dis)similarities between them. Fourth, teachers found center emojis (i.e., centroids) helpful to represent information in a concise way for more effective pattern recognition. Last, teachers confirmed the potential of applying cluster analysis as a discovery tool as it supports students to identify different types of patterns (e.g., intra & inter-cluster pattern, centroid) for further interpretation.

Patterns interpreted for scientific hypothesis formulation

While formulating a scientific hypothesis in the end, we found that teachers interpreted different patterns recognized from cluster analysis in different ways. First, teachers utilized patterns of inter-cluster differentiating features in three different approaches: (1) two teachers' attention was caught by the cluster with the minimum or maximum value of some data attribute; for example, P1 identified the cluster of field sites that has the highest precipitation, the highest temperature, and the highest beetle richness as evidence for their hypothesis; (2) nine teachers interpreted differentiating features between different pairs of clusters; for example, P12 recognized that, compared to the first field site cluster, the second cluster of field sites are drier, have more beetles and higher canopy, and the third cluster has more beetles, and the fourth has lower latitude, more beetles, higher temperature and higher precipitation; (3) five out of these nine teachers successfully formed hypotheses on relationships among ecological factors by further interpreting the identified inter-cluster differentiating features; for example, P9 hypothesized that "latitude and temperature have a large impact on the beetle richness and canopy height", while P12 was not able to produce a more meaningful hypothesis with eight pairs of inter-cluster differentiating features identified.

Second, three teachers paid attention to inter-cluster shared features. P11 identified canopy height, precipitation, and latitude as the similarities shared among the field sites of interest. Third, while 11 teachers considered intra-cluster similarities, only P2 paid attention to intra-cluster variations: "small mammal richness varies widely even in a cluster". In addition, another very unique pattern utilized is that P13 compared the similarities between each pair of clusters and identified the cluster that is the most dissimilar to the rest clusters, as the most distinct cluster and identified the cluster it is most distinct from.

RQ2: What challenges prevent teachers from conducting effective scientific inquiry?

Cognitive challenges for novice learners to identify multidimensional patterns from face-based glyphs

By analyzing feedback and log data around pattern recognition with face-based glyphs, we identified cognitive challenges related to (1) concept mapping and (2) value mapping in face-based glyph visualization.

Related to the concept mapping where data attributes are mapped to facial features, we identified different causes introducing extra cognitive load into pattern recognition. First, the legend showing corresponding data attributes when the cursor is hovering over a certain area is still temporarily visible, which leads to the lack of connection to scientific context. Second, teachers had attention bias that they perceived certain facial features as more conspicuous than others. Third, from the patterns teachers identified from glyphs, we recognized inaccurate pattern recognition caused by ambiguous or less accurate verbalization of certain visual features due to people's habits in language using while describing faces. For example, six teachers described mouth features with inaccurate language, such as using mouth shape to refer to mouth curvature which caused confusion in the interpretation of mouth size and mouth curvature as both affect mouth shape.

With the value mapping of face-based glyphs where different data attributes' values are translated into different visual channels, we recognized the difficulty in deciding if certain values are (dis)similar enough because of the low separability of the multiple values mapped to one face part (e.g., brow length, brow vertical location, brow slant are all mapped to eyebrows). Teachers also commented that it's hard for them to decide if a certain pattern should be identified as a large/medium/small value or if certain degrees of variations should be tolerant because there is no visual indicator for the dispersion of each facial feature. Furthermore, by analyzing the patterns identified during cluster analysis, we identified that teachers have different tolerance for variations in different facial features. On the one hand, teachers discerned deviations in mouth vertical, mouth curve, and eye radius, with 17.5%, 15.63%, and 13.89% instances of high tolerance for variations in hypothesis development, respectively. On the other hand, teachers tended to be more sensitive to changes in eyebrow length (16.67%), eyebrow slant (11.36%), and nose length (11.11%).

Challenges in more sufficient pattern interpretation during hypothesis generation

By measuring the quality of teachers' pattern interpretation for evidence-based hypothesis generation in the end, we identified teachers who (1) failed to illustrate the patterns, (2) needed further inference, (3) needed to analyze the direction of the correlation, (4) needed to have a contrast group to make the inference. We found two main causes of deficient hypothesis generation:(1) the insufficient interpretation and synthesis of different patterns, (2) lacking in-depth analysis for more sophisticated multidimensional data inference.

First, five out of six teachers who hypothesized relationships among ecological factors didn't further elaborate on whether the relationships are positive or negative by interpreting the value mappings in the patterns identified. For example, P14 interpreted that clusters with different latitudes have different amounts of precipitation, instead of elaborating that the field sites with higher latitudes share lower amounts of precipitation than the cluster with lower latitudes. Besides such insufficient interpretation of data attributes' values, we observed that five teachers lacked synthesizing multiple pattern interpretations to formulate a more meaningful scientific hypothesis. Such insufficient synthesis could indicate the need for some basic domain knowledge about ecosystems to guide learners' attention to what patterns could be more interesting to focus on or how the information behind different patterns can be effectively synthesized and form a scientific hypothesis leading to a more in-depth understanding of the domain. By looking into the interaction data of two teachers who failed to illustrate the patterns from cluster analysis, we found that one of them only compared two centroids, and the other only examined one cluster centroid. Both of them didn't explicitly identify the value mapping during pattern recognition. This further indicates that the insufficient interpretation and synthesis of different patterns may lead to unsuccessful evidence-based hypothesis generation and corresponding scaffoldings are needed.

Second, some teachers lacked more careful analysis of different patterns and missed more sophisticated inferences on different ecological attributes. For example, P4 hypothesized that canopy height influences beetle richness. Without digging into the other shared or differentiating features among clusters, P4 didn't interpret that only for field sites with medium amounts of precipitation, canopy height and beetle richness are positively related. Therefore, P4 missed introducing precipitation as a moderator to form a more elaborated data-driven scientific hypothesis. We also observed the case where teachers didn't include multiple contributing features as independent variables for the dependent variables in their hypotheses, and the case where a teacher didn't more carefully support their hypothesis with comparison from patterns identified.

Discussion

For each challenge identified from novice learners' data recognition and interpretation in clustering-supported scientific hypothesis generation, we summarized a set of fundamental learner needs underneath all the challenges we identified from novice learners' pattern recognition and interpretation for generating clustering-supported hypotheses with face-based glyphs (Table 2). **First**, learners need more in-time access to the scientific context and more effective scientific communication with smoother code-switching between glyph visual features and data attributes (N1.1, N1.2). Therefore, more design opportunities lie in the embedded legend design for glyph visualization with continuous visibility or an intelligent mapping mechanism creating concept mapping with better learnability (Chung et al., 2015). **Second**, during identifying multidimensional patterns from glyphs, learners are viewing all data attributes at once and thus need scientifically guided attention to select features of interest and interpret the visual channel in a more informed and accurate way (N1.3, N2.1). This suggests a potential design of an intelligent value mapping mechanism utilizing people's varying attention across facial features and visually related glyph features (Chung et al., 2015). **Third**, when digging into the interpretation of a specific data attribute, learners need more measurable displays of that attribute's spread and deviation (N2.2, N2.3). Thus, more in-time indicators of a comprehensive list of parameters describing the distribution of each data attribute (e.g., center, spread, shape) are needed for glyph visualization. **Fourth**, various multidimensional patterns can be identified and interpreted for cluster analysis, which is challenging for novices to analyze, interpret and evaluate as different forms of evidence in relation to potential scientific hypotheses (Kuhn & Lerman, 2021) (N3.1, N3.2). This could be scaffolded via probabilistic reasoning and contrastive explanation in the context of ML-enhanced scientific inquiry (Miller, 2019). For example, when commonalities between two clusters are observed, then learners can be reminded of what patterns are missing as comparisons in order to indicate correlations between data attributes.

Table 2
Design opportunities of face-based glyph visualization for multidimensional data interpretation

Challenges	Underlying fundamental needs as design opportunities
C1.1 Extra cognitive load from the temporarily visible concept mapping	***N1.1*** Easy access to the scientific context behind the data
C1.2 Less precise verbalization	***N1.2*** Effective scientific communication of multivariate data and patterns
C1.3 Attention bias across facial features	***N1.3*** Guided attention for more informed and effective selection of features of interest
C2.1. Low separability of some visual features ***C2.2*** Absence of the scale	***N2.1*** Independent interpretation of each visual channel ***N2.2*** Spread of each data attribute
C2.3 Uncertainty in variation interpretation	***N2.3*** Deviation of each data attribute
C3.1 Insufficient interpretation and synthesis of different patterns	***N3.1*** Guided analysis and synthesis of all the potential patterns and their relations
C3.2 Lacking in-depth analysis for sophisticated multivariate inference	***N3.2*** Evaluation of different patterns in relation to intended inferences

As the initial stage in an iterative design of data visualization and interaction for ML-enhanced K-12 STEM learning, we only set up a preliminary framework connecting cluster analysis with the conceptualization in inquiry-based learning and haven't fully implemented and evaluated it with both K-12 teachers and students. Additionally, we noticed that the step-by-step scaffolding for cluster analysis in the current learning activity design could guide more learner attention to intra-cluster shared features and inter-cluster differentiating features. Therefore, as a next step, we are planning to create new scaffoldings for a more open-ended learning environment to empower students to explore the data and create scientific questions or hypotheses. Besides collecting feedback from teachers with student perspectives, we will evaluate new designs with young students.

Conclusion

This work proposes a framework showing the pedagogical and cognitive potential of using face-based glyphs to introduce cluster analysis as an accessible and versatile tool for K-12 inquiry-based learning. Eighteen K-12 STEM teachers conducted scientific inquiry through cluster analysis. We evaluate their performance under pattern recognition and interpretation for hypothesis generation. Challenges are identified in recognizing patterns with face-based glyphs, conceptually and extensively synthesizing different types of patterns, and scrutinizing for more sophisticated multidimensional relationships. We reflect on the fundamental learner needs behind these challenges and discuss corresponding design opportunities and implications.

References

Anderberg, M. R. (2014). Cluster Analysis for Applications: Probability and Mathematical Statistics: A Series of Monographs and Textbooks. Academic Press.

Braham, H. M., & Ben-Zvi, D. (2017). STUDENTS' EMERGENT ARTICULATIONS OF STATISTICAL MODELS AND MODELING IN MAKING INFORMAL STATISTICAL INFERENCES. In STATISTICS EDUCATION RESEARCH JOURNAL (Vol. 16, Issue 2, pp. 116–143). https://doi.org/10.52041/serj.v16i2.187

Cao, N., Gotz, D., Sun, J., & Qu, H. (2011). DICON: interactive visual analysis of multidimensional clusters. IEEE Transactions on Visualization and Computer Graphics, 17(12), 2581–2590.

Chernoff, H. (1973). The Use of Faces to Represent Points in k-Dimensional Space Graphically. Journal of the American Statistical Association, 68(342), 361–368.

Chin-Parker, S., & Bradner, A. (2017). A contrastive account of explanation generation. Psychonomic Bulletin & Review, 24(5), 1387–1397.

Chung, D. H. S., Legg, P. A., Parry, M. L., Bown, R., Griffiths, I. W., Laramee, R. S., & Chen, M. (2015). Glyph sorting: Interactive visualization for multi-dimensional data. Information Visualization, 14(1), 76–90.

Folger, R., & Stein, C. (2017). Abduction 101: Reasoning processes to aid discovery. Human Resource Management Review, 27(2), 306–315.

Greenwald, E., Leitner, M., & Wang, N. (2021). Learning Artificial Intelligence: Insights into How Youth Encounter and Build Understanding of AI Concepts. Proceedings of the AAAI Conference on Artificial

Intelligence, 35, 15526–15533.

Han, J., Pei, J., & Kamber, M. (2011). Data Mining: Concepts and Techniques. Elsevier.

Jong, T. de, de Jong, T., Sotiriou, S., & Gillet, D. (2014). Innovations in STEM education: the Go-Lab federation of online labs. In Smart Learning Environments (Vol. 1, Issue 1). https://doi.org/10.1186/s40561-014-0003-6

Kazak, S., Fujita, T., & Turmo, M. P. (2021). Students' informal statistical inferences through data modeling with a large multivariate dataset. Mathematical Thinking and Learning, 1–21.

Keck, M., Kammer, D., Gründer, T., Thom, T., Kleinsteuber, M., Maasch, A., & Groh, R. (2017). Towards Glyph-based visualizations for big data clustering. Proceedings of the 10th International Symposium on Visual Information Communication and Interaction, 129–136.

Kitano, H. (2016). Artificial Intelligence to Win the Nobel Prize and Beyond: Creating the Engine for Scientific Discovery. AI Magazine, 37(1), 39–49.

Koehler, M., & Mishra, P. (2009). What is technological pedagogical content knowledge (TPACK)? Contemporary Issues in Technology and Teacher Education, 9(1), 60–70.

Kuhn, D. (2016). What do young science students need to learn about variables? Science Education, 100(2), 392–403.

Kuhn, D., & Lerman, D. (2021). Yes but: developing a critical stance toward evidence. International Journal of Science Education, 1–18.

Kuhn, D., Ramsey, S., & Arvidsson, T. S. (2015). Developing multivariable thinkers. Cognitive Development, 35, 92–110.

Lee, M. D., Reilly, R. E., & Butavicius, M. E. (2003). An empirical evaluation of Chernoff faces, star glyphs, and spatial visualizations for binary data. Proceedings of the Asia-Pacific Symposium on Information Visualisation-Volume 24, 1–10.

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. Artificial Intelligence, 267, 1–38.

Morris, C. J., Ebert, D. S., & Rheingans, P. L. (2000). Experimental analysis of the effectiveness of features in Chernoff faces. In 28th AIPR Workshop: 3D Visualization for Data Exploration and Decision Making. https://doi.org/10.1117/12.384865

Muller, M., Guha, S., Baumer, E. P. S., Mimno, D., & Shami, N. S. (2016). Machine Learning and Grounded Theory Method: Convergence, Divergence, and Combination. Proceedings of the 19th International Conference on Supporting Group Work, 3–8.

Nayak, J. G., Hartzler, A. L., Macleod, L. C., Izard, J. P., Dalkin, B. M., & Gore, J. L. (2016). Relevance of graph literacy in the development of patient-centered communication tools. Patient Education and Counseling, 99(3), 448–454.

Pedaste, M., Mäeots, M., Siiman, L. A., de Jong, T., van Riesen, S. A. N., Kamp, E. T., Manoli, C. C., Zacharia, Z. C., & Tsourlidaki, E. (2015). Phases of inquiry-based learning: Definitions and the inquiry cycle. In Educational Research Review (Vol. 14, pp. 47–61). https://doi.org/10.1016/j.edurev.2015.02.003

Rau, M. A. (2017). Conditions for the effectiveness of multiple visual representations in enhancing STEM learning. Educational Psychology Review, 29(4), 717–761.

Rodrigues, C. T. (2011). The Method of Scientific Discovery in Peirce's Philosophy: Deduction, Induction, and Abduction. In Logica Universalis (Vol. 5, Issue 1, pp. 127–164). https://doi.org/10.1007/s11787-011-0026-5

Romesburg, C. (2004). Cluster Analysis for Researchers. Lulu.com.

Roth, W.-M., Michael Bowen, G., & McGinn, M. K. (1999). Differences in graph-related practices between high school biology textbooks and scientific ecology journals. In Journal of Research in Science Teaching (Vol. 36, Issue 9, pp. 977–1019). https://doi.org/10.1002/(sici)1098-2736(199911)36:9<977::aid-tea3>3.0.co;2-v

States, N. L. (2013). Next Generation Science Standards: For States, By States. National Academies Press.

Strauss, A., & Corbin, J. (1990). Basics of qualitative research. Sage publications.

Wan, X., Zhou, X., Ye, Z., Mortensen, C. K., & Bai, Z. (2020). SmileyCluster: supporting accessible machine learning in K-12 scientific discovery. Proceedings of the Interaction Design and Children Conference, 23–35.

Ward, M. O., Grinstein, G., & Keim, D. (2015). Interactive Data Visualization: Foundations, Techniques, and Applications, Second Edition. CRC Press.

Zhou, X., Tang, J., Daley, M., Ahmad, S., & Bai, Z. (2021). "Now, I Want to Teach It for Real!": Introducing Machine Learning as a Scientific Discovery Tool for K-12 Teachers. Artificial Intelligence in Education, 486–499.