Creating an Authoring Tool for K-12 Teachers to Design ML-Supported Scientific Inquiry Learning

Xiaofei Zhou
xzhou50@ur.rochester.edu
University of Rochester
Rochester, New York, USA

Jingwan Tang
jtang21@u.rochester.edu
University of Rochester
Rochester, New York, USA

Hanjia Lyu
hlyu5@ur.rochester.edu
University of Rochester
Rochester, New York, USA

Xinyi Liu
xinyi.liu@utexas.edu
The University of Texas at Austin
Austin, Texas, USA

Zhenhao Zhang
zzh126@ur.rochester.edu
University of Rochester
Rochester, New York, USA

Lichen Qin
lqin5@ur.rochester.edu
University of Rochester
Rochester, New York, USA

Fiona Au
fau@u.rochester.edu
University of Rochester
Rochester, New York, USA

Advait Sarkar
advait@microsoft.com
Microsoft Research
Cambridge, United Kingdom

Zhen Bai
zbai@cs.rochester.edu
University of Rochester
Rochester, New York, USA

ABSTRACT

Despite significant advances in machine learning (ML) applications within science, there is a notable gap in its integration into K-12 education to enhance data literacy and scientific inquiry (SI) skills. To address this gap, we enable K-12 teachers with limited technical expertise to apply ML for pattern discovery and explore how ML can empower educators in teaching SI. We design a web-based tool, ML4SI, for teachers to create ML-supported SI learning activities. This tool can also facilitate collecting data about the interaction between ML techniques and SI learning. A pilot study with three K-12 teachers provides insights to prepare the next generation for the era of big data through ML-supported SI learning.

CCS CONCEPTS

• **Applied computing → Interactive learning environments**; • **Social and professional topics → Computing education**; • **Human-centered computing → Empirical studies in HCI**.

KEYWORDS

Machine learning, scientific inquiry, K-12 STEM education, AI education, data visualization, end-user programming

ACM Reference Format:
Xiaofei Zhou, Jingwan Tang, Hanjia Lyu, Xinyi Liu, Zhenhao Zhang, Lichen Qin, Fiona Au, Advait Sarkar, and Zhen Bai. 2024. Creating an Authoring Tool for K-12 Teachers to Design ML-Supported Scientific Inquiry Learning. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '24), May 11–16, 2024, Honolulu, HI, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3613905.3650762

CHI EA '24, May 11–16, 2024, Honolulu, HI, USA

ACM ISBN 979-8-4007-0331-7/24/05
https://doi.org/10.1145/3613905.3650762

1 INTRODUCTION

Machine Learning (ML) has become a powerful tool for scientific discovery, enabling new problem-solving approaches in several scientific disciplines [2, 24, 35, 37]. Techniques such as clustering and classification uncover patterns in large datasets that were previously intractable, accelerating data-driven knowledge discovery [6, 16, 51]. Scientific inquiry (SI) learning advocates for K-12 students to adopt practices akin to professional scientists for knowledge construction [25]. In the era of big data and ML, it is therefore important to explore the design of ML-supported SI activities for K-12 STEM education [34]. The aspiration is to cultivate authentic data and computational literacy at an early age, thereby equipping the next generation with the expertise needed in a data-centric world [49].

There remains a gap between ML concepts and K-12 STEM pedagogy [20, 45, 57]. To bridge the gap with meaningful learning experiences for students from diverse backgrounds, it is crucial to involve K-12 teachers in creating learning activities [30, 38, 54]. Teachers, however, often have limited ML expertise [30, 44] to craft pedagogically sound ML practices that captivate students' interests [31, 45]. To address these challenges, our ongoing work makes two contributions:

(1) We present **ML4SI**, a web-based tool for K-12 teachers to design ML-supported SI learning activities. It allows teachers to construct learning activities by arranging pre-designed ML & SI components in a side-by-side layout.
(2) With the data collected from ML4SI, we can model **how K-12 teachers and students apply different ML methods along with various SI behaviors**, such as questioning and formulating hypotheses [34].

Preliminary findings from a pilot study with three K-12 teachers suggest strengths and opportunities for design for ML4SI, and a potential interplay between ML techniques and SI learning behaviors.

2 RELATED WORK

2.1 Design Guidelines for End-User Programming Environments

Various guidelines and frameworks have been proposed for designing programming environments to support non-expert end-users [4]. The cognitive dimensions of notations [18], for instance, provide a vocabulary by which notation designers can discuss the trade-offs made by design choices. These dimensions provide insights into the user interface's navigability, consistency, and error-proneness, making them a valuable tool for assessing and improving the design of systems intended for non-programmers. Building upon these dimensions, Repenning and Ioannidou's 13 design guidelines are more implementation-focused [36]. For instance, strategies like minimizing the possibility of syntactic errors, incorporating objects as language elements, and fostering incremental development are directly applicable during the design phase, ensuring a more accessible and intuitive user experience. Sarkar [39] proposes four design principles for non-expert data analysis through ML: (1) start the abstraction gradient at zero, (2) abstract complex processes through heuristic automation, (3) build expertise through iteration on multiple representations, and (4) support dialogue through meta-models. These design dimensions and guidelines offer insights into the ML4SI interface and our pilot study with K-12 teachers.

2.2 Inquiry-Based Learning

Inquiry-based learning serves as an educational approach designed to engage students in authentic SI practices [25]. It emphasizes students' active engagement in the SI learning process and their responsibility for uncovering novel knowledge through a dual process of inductive and deductive reasoning [9, 50].

In ML4SI, we adopt a well-recognized framework for inquiry-based learning [34] to extract the essential SI learning behaviors. From the five key SI learning phases, we primarily focus on conceptualization, investigation, and conclusion—since they have a direct connection with data.

Conceptualization illustrates the process where students actively propose research questions to be explored or hypotheses to be tested. *Investigation* is where students investigate research questions or hypotheses through exploration, experimentation, analysis, and data interpretation—exploration emphasizes exploratory data analysis and observation in which students make discoveries related to their questions without a predetermined hypothesis; students, during experimentation, test specific hypotheses by designing and executing experiments; analysis is systematically analyzing data to identify patterns and draw meaningful inferences; data interpretation derives meanings from patterns revealed by data analysis. *Conclusion* is where students address their original research questions or hypotheses and determine whether they have been effectively answered or supported by the results obtained from their study.

2.3 ML-Supported Scientific Inquiry

ML techniques, known for their ability to learn from experience and identify complex patterns within data, have been rapidly adopted by scientists across various disciplines in the pursuit of SI [1–3, 35, 37]. ML equips scientists with a powerful toolkit for automating the analysis of large datasets, accelerating systematic investigations and explorations [16, 32], and uncovering novel patterns that are often hidden from conventional techniques [5]. Moreover, it can guide the design of experiments and shape future data collection strategies [27].

Despite the demonstrated potential of ML in SI, its integration into K-12 STEM education has been slow. Two pioneering studies shed light on this matter. One study observed early signs of SI learning behaviors, such as questioning and explaining, from students during a hands-on data analysis experience that teaches K-means clustering [46]. Another study used a predefined SI model to guide K-12 teachers to create conceptual designs of ML-supported SI learning activities [44, 55]. However, the structured nature of the predefined model hindered the fidelity of the designs in real-world teaching scenarios, compared with an authentic ML tool.

To address this research gap, we develop ML4SI, an authoring tool that provides teachers with flexibility in integrating ML into SI practices, focusing on two ML algorithms: K-means clustering and K-nearest neighbors (KNN). K-means clustering, an unsupervised ML algorithm, is employed to discover patterns among similar objects and create taxonomies [13, 14, 47], making it suitable for exploratory data analysis [5]. K-nearest neighbors, a supervised ML algorithm for the classification task, with its primary objective being to predict the class of a data point based on known examples.

2.4 Multidimensional Data Visualization

The interpretability of ML-revealed patterns is a significant concern, particularly for K-12 teachers and students with limited skills in reading and interpreting multidimensional data [14, 29, 41].

There are two main types of multidimensional data visualization: geometric methods and iconographic displays [12]. *Geometric methods* portray multidimensional data using the axes of selected shapes [12]. Parallel coordinates, a widely-used geometric method [12, 22], represent attributes by parallel vertical axes linearly scaled within respective data ranges and depict each data point by a polygonal line that intersects each axis at the corresponding value. *Iconographic displays*, also known as glyphs, encode attributes and values into visual features, such as size, shape, and color [12, 48]. It makes multidimensional information more accessible, particularly for novice learners, compared with geometric methods [12, 48]. Face glyphs, star glyphs, and profile glyphs are three common glyph methods [7, 11, 12]. Face glyphs map attributes and values to facial features, such as eye size and nose width [8]. Star glyphs represent attributes as spokes of a circular wheel, with values encoded by the length from outer points to a central point [15, 48]. Profile glyphs use linear position/length to encode values [7, 11].

Each aforementioned visualization technique offers distinct advantages for multidimensional data analysis. For example, parallel coordinates are particularly effective for discerning data distribution and functional dependencies [52]. Face glyphs leverage human familiarity with facial elements, facilitating more effective data integration and stimulating greater interest in engaging with data analysis [8, 23, 28]. Star glyphs and profile glyphs resemble conventional visualization tools familiar to K-12 students, such as pie graphs and bar charts. We incorporate different forms of visualizations into ML4SI, allowing teachers to decide the most suitable visualization for specific inquiry activities based on their expertise.

3 ML4SI

3.1 Design Guidelines

We leverage design implications derived from a co-design study with K-12 teachers [44, 56] and existing design guidelines for end-user programming environments [4, 36] for developing ML4SI.

First, *"abstraction gradient"* assesses how the tool abstracts complex ML processes into user-friendly interfaces. ML4SI simplifies ML algorithms like K-means clustering into drag-and-drop components (Section 3.2.1), making advanced data analysis accessible to teachers with limited ML expertise. Second, *"closeness of mapping"* evaluates how well the tool's interface corresponds to the real-world tasks of K-12 teachers. Our system's side-by-side layout of ML and SI components (Section 3.2.2) mirrors the natural workflow of teachers planning and executing lessons, thus aligning closely with their educational objectives and problem-solving domain. Third, to support non-experts' *incremental development* of ML-supported analysis, ML4SI enables teachers to easily customize input for individual ML components (Section 3.2.4) and execute before the completion of the entire process. Such step-by-step creation with ML components that gradually reveal the computational complexity (Section 3.2.1) avoids huge leaps in the challenge for novices. Finally, ML4SI *allows immersion.* Teachers are immersed in ML-supported tasks and actively experience the results through directly dragging & dropping data features onto glyph visual features (Fig. 1.2, left column) along with real-time updates of results in the consecutive steps (Section 3.2.5). Glyph visualizations (Section 3.2.3) also have unique learning benefits for end users by affording direct manipulation and interaction with multidimensional data points.

3.2 Design Features, By Example

Jeremy, a middle-school biology teacher, aims to demonstrate the dynamic interactions between various ecological features. His objective is to create an SI learning activity for students to uncover relationships among temperature, precipitation, canopy height, mammal richness, and beetle richness. This task involves applying ML algorithms to a multidimensional dataset gathered from extensive field sites. Despite his high motivation, Jeremy has not yet acquired sufficient expertise in ML to construct an effective lesson for his students. ML4SI is designed to assist teachers like Jeremy in realizing such educational objectives.

3.2.1 Drag & Drop Blocks to Initiate ML/SI Components. The top bar of ML4SI has draggable components (Fig. 1.1). When Jeremy drags and drops ML/SI components into the main workspace (Fig. 1.2), the corresponding ML methods or SI behaviors are initiated. This enables Jeremy to begin exploring the dataset, interpreting patterns, and creating specific steps in the learning activity.

ML components (Fig. 1.1, left column) allow the application of ML techniques. For example, the pairwise comparison component computes the similarity between two data points. Manual clustering [46] reveals patterns in a subset by allowing users to manually overlay glyphs for similarity comparison through superposition comparative visualization [17]. Automatic clustering applies k-means clustering on input data and visualizes the clusters. Prediction with classifiers guides users to predict unlabeled data with KNN and evaluate the prediction result.

SI components (Fig. 1.1, right column) represent established SI learning behaviors (e.g., questioning, hypothesis generation, data analysis, conclusions) [34] as introduced in Section 2.2. Since inquiry-based learning is self-directed, students may encounter challenges in the learning process if they lack the necessary self-regulation skills [53]. ML4SI therefore includes an instruction component that teachers can use to create scaffolds to keep students metacognitively, motivationally, and behaviorally engaged.

3.2.2 Side-by-Side Layout. This layout reflects the visual presentation of the final design (Fig. 1.2). Users are able to add, reorder, and remove individual or pairs of ML and SI components to tailor the learning design. Jeremy experiments with different pairs of ML and SI components, by considering which ML components would best support a particular SI learning behavior, or which SI behaviors students are most likely to demonstrate while analyzing with a specific ML technique.

3.2.3 Data Visualization Selection for Individual ML Components. Jeremy can select from multiple data visualizations for each ML component (Fig. 1.3). The choice of visualization in one component does not affect the others, allowing for independent selection across different inquiry phases. More visualization options can be added for different ML techniques.

For instance, Jeremy might opt for face glyphs in early data exploration to spark students' interest in the intriguing patterns revealed by the "smileys". Later, he could switch to parallel coordinates for his students to examine more detailed numerical patterns.

3.2.4 Input and Output for Individual ML Components. In each ML component, Jeremy can input different data, ranging from the entire dataset to a cluster formed by a previous clustering component, or a manually selected subset (Fig. 2.1 & 2.4). Additionally, Jeremy can save the outputs of ML components. For instance, he might preserve the clusters created by K-means clustering (Fig. 2.2 & 2.3) for another round of clustering.

3.2.5 Real-Time Updates for Feature Selection. In feature selection (Fig. 1.2, left column), any edits also trigger updates of subsequent ML components. For example, Jeremy creates parallel coordinates using all five data features from a five-dimensional dataset, followed by automatic clustering. Clusters will then be generated out of the five-dimensional data. Through cluster analysis, Jeremy discovers that there are still large variations in the last data feature "mammal richness" in all clusters (Fig. 2.2). He hypothesizes that this is not an important feature for clustering. He tests this by removing the feature in the feature selection, and the change is immediately reflected in the clustering results. This allows users to refine their hypotheses and explore the impact of feature selection on ML outputs, encouraging experimentation and trial & error.

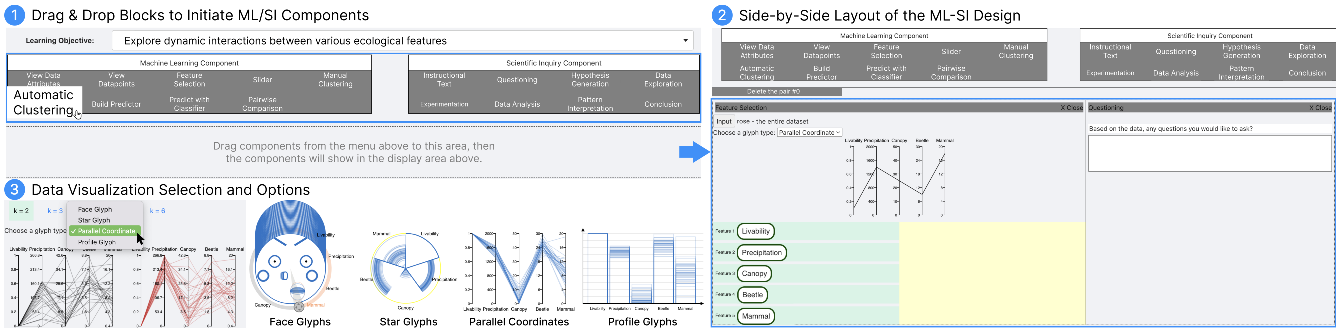

Figure 1: (1) Draggable blocks to initiate ML & SI components; (2) Side-by-side layout for adding and designing ML (e.g., feature selection) and SI components (e.g., questioning); (3) A drop-down list to select a visualization method for a specific ML component from the options provided.

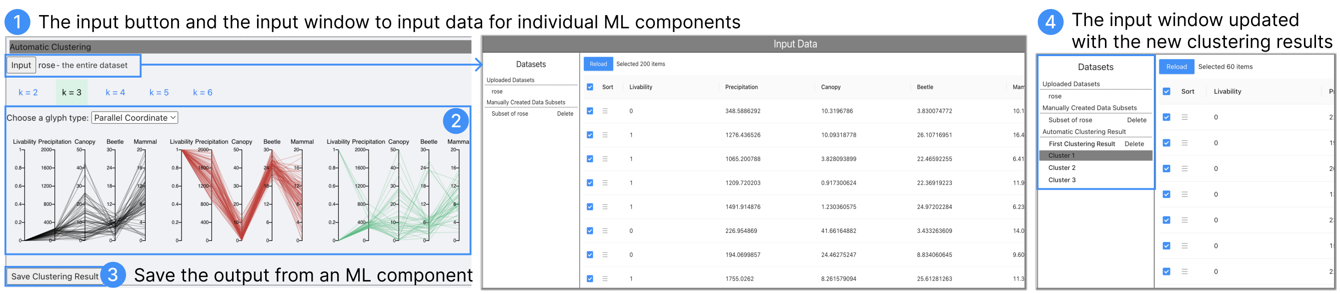

Figure 2: (1) Select the entire dataset as the input for k-means clustering; (2) visualize the clustering results with parallel coordinates; (3) click the button to save the clusters; (4) the input window is updated after the new clusters are saved from k-means clustering.

4 PILOT STUDY

4.1 Participants

As an initial investigation, we invited three K-12 STEM teachers (Table 1) to try ML4SI. We wished to evaluate ML4SI with K-12 teachers in a co-design setting for three reasons: (1) teachers' expertise is crucial due to the limited pedagogical theories on integrating ML methods into K-12 scientific inquiry [44, 45]; (2) teachers often underestimate their ability to teach with ML, leading to low self-efficacy; adequate support can mitigate this issue [21, 26, 30, 43]; and (3) involving teachers and ML experts helps balance between user-centered and learner-centered methodologies; teachers can engage with complex data and uncover meaningful patterns with the support of experts [10, 19].

4.2 Study Procedure

The pilot study consists of two two-hour sessions. The study was approved by the institutional Research Subjects Review Board.

Session 1: Teacher-as-Learner. Teachers familiarised themselves with the ML4SI interface by analyzing an example dataset under researcher guidance, including (1) different data visualizations, (2) ML components, (3) SI components, and (4) the flow of designing the learning activity in ML4SI. Participants asked questions and provided feedback after interacting with each ML/SI component. At the end of the first session, the teacher and the researcher discussed the types of datasets or learning activities the teacher wished to create during the second session.

A minimum of two days separated the first and second sessions. In this interval, either the researcher or the participant acquired a new dataset as per the discussion in the first session. The researcher pre-processed the new dataset by eliminating data points with missing values and non-numeric features. This ensured the dataset's compatibility with the ML algorithms implemented in ML4SI.

Session 2: Teacher-as-Designer. Teachers explored the dataset of their choice with ML4SI. A researcher and an ML expert provided facilitation when needed. Teachers asked questions that interested them at each step, answered the inquiries based on the data visualization, and planned subsequent actions. Following each step, the researcher prompted teachers to consider any necessary adjustments to the pairing of ML and SI components. Upon completion, teachers were asked to review their process and contemplate modifications to the learning activity they created. Finally, teachers provided feedback on the system design.

4.3 Initial Data Collection and Analysis

We collect video recordings of the study sessions and the log data of teachers' interaction behaviors with ML4SI. We had two aims. First, to investigate the analysis process of teachers applying ML methods to explore the datasets of their interest. Second, to understand how

Table 1: Demographics of study participants.

PID	Gender	Teaching Grade(s)	Teaching Subject(s)	Years of Teaching
P1	Female	High school	Algebra and AP Statistics	16 years
P2	Female	Middle & high school	Computer Science, Business Technology	5 years
P3	Male	High school	Math, Science	8 years

different ML components support different SI behaviors in the final designed learning activity.

5 PRELIMINARY RESULTS

5.1 An Example Lesson Created in Our Pilot

This section describes a lesson designed by P2, a CS teacher in a city school where students have low digital literacy and are from underrepresented backgrounds in STEM. The learning objective is to investigate what impacts a person's income level. The essential question for students to answer in the end is how college and career planning in high school affect your life.

First, P2 began with introducing students to the key concepts involved in the dataset by viewing data attributes' definitions. *Second*, P2 guided students to select specific data attributes - working hours, age, gender, and education level - aligning with the target learning objective. *Third*, P2 asked students to compare income predictions, aiming to inspire students to question: "why some individuals are predicted to have high income and others low?" After applying KNN a few times, she hypothesized positive correlations between higher education, longer working hours, and higher income. P2 expected her students would be able to achieve the same SI learning behavior while interacting with KNN. *Fourth*, P2 directed students to validate initial hypotheses with the patterns revealed by applying k-means clustering on the entire dataset. P2 discovered trends that clusters with high incomes have higher education, longer working hours, and medium ages, while the low-income cluster has lower education and fewer working hours (Fig. 3). *Lastly*, P2 added a step for students to answer the essential question by contextualizing the patterns in social, economic, and literacy development domains.

5.2 Feedback from K-12 STEM Teachers

A structured and flexible way to create SI learning activities. All teachers appreciated how the side-by-side layout affords a structured way of conducting the open-ended exploration and iterating the SI learning design after reflection. Edit-triggered updates of feature selection and the ability to investigate any data subset helped teachers experiment with ideas, and zoom in and out on different parts of the data. This enabled a deeper understanding of the patterns and the ML algorithms through trial & error. After reviewing their analysis process, teachers tended to create the final design to incrementally build up the complexity of hypotheses, in an attempt to scaffold for students to apprehend the challenges of iterating hypotheses with different analysis methods applied to different data. Furthermore, teachers recognized ML4SI's adaptability to incorporate existing teaching techniques, such as the graphic organizer, the Three-Two-One technique (i.e., identify three interesting observations, ask two questions, raise one potential solution), self-directed learning for small group activities, etc.

Integrating ML literacy into K-12 STEM contexts. All teachers noted close connections between ML practices to K-12 science objectives, data/computational literacy, and digital fluency required by K-12 curriculum standards [33, 40, 42]. For example, comparing the classification results with the ground truth and tinkering with the ML parameter supports the development of debugging skills. Visualizations can make a range of patterns accessible for children to analyze, such as intra-cluster patterns, outliers, variations, etc. One math teacher pointed out that analyzing centroid, intra-cluster similarity, and variation can create the opportunity for his students to discuss the issues of using the average to represent a group of data points with a large variation. Participants suggested a range of K-12 STEM topics that can be taught with ML4SI, including career readiness, nutrition, college applications, data-driven strategy for football drafting, digital citizenship, bias and privacy issues in data, AI recommendation systems, etc.

Introducing automated assistance. We noted opportunities for adding intelligent assistance. One teacher suggested that students' hypothesis generation and iteration can be automatically identified, extracted, and tracked throughout the analysis, and thus, scaffolding can be personalized for different learners at different SI phases. After manually selecting a few data points of his interest as the input for manual clustering, one teacher suggested having the system suggest input for an ML component to generate the optimal or insightful results, based on the existing steps.

Opening the "black box". Teachers asked for more technical details underlying the clustering and classification algorithms, and more advanced statistical information about the data. For example, beyond learning about how KNN predicts an unknown value for a data point based on its nearest neighbors, one teacher was also curious about how to improve the labeled data to increase prediction accuracy. Another teacher requested a more advanced and detailed statistical summary for a specific cluster.

Benefits and Limitations of Glyphs. Glyphs encode data at the level of individual data points, as opposed to chart types that aggregate data (e.g., bar charts). Teachers identified glyphs' unique benefit in directly visualizing individual data points for students to manipulate and reason with, such as (1) the representation and calculation of centroid, (2) the relationship between average, variation, and individual values in a data group, (3) how global patterns such as trends and correlations emerge from individual data points, and (4) the algorithmic mechanism underlying KNN. This builds upon Sarkar's design principle of directly representing individual data points for non-experts doing data analysis through ML, i.e., starting the abstraction gradient at "zero" [39]. With the aforementioned benefits of using glyphs for data exploration in K-12 classrooms, it is important to note that the scalability of glyphs is limited by

Figure 3: The pair of the ML and SI components (automatic clustering and data analysis) designed by P2 for high-school students to investigate factors that impact a person's income level.

the number of data features and the quantity of data points to be visualized.

Affordances of different visualizations. Teachers appreciated the ability to choose from multiple visualization types. Two teachers whose students have lower data and computational literacy preferred face glyphs, compared to the traditional graphs introduced in K-12 classrooms. They suggested that face glyphs could engage students who are not proficient in reading data graphs. Teachers emphasized that if there are positive or negative implications behind some data attributes, face glyphs can establish semantic meaning between the visual representation and the patterns, and thus promote students' analysis. One teacher envisioned that his students might tell a compelling story about how to win baseball games by interpreting the patterns visualized in happy faces.

6 CONCLUSIONS AND FUTURE WORK

The preliminary findings indicate that the integration of ML into K-12 education can serve as an opportunity to enhance data literacy and SI skills for young students. ML4SI is a step towards bridging the gap between advanced ML tools and their use in K-12 STEM classrooms, and demystifying ML. Furthermore, the positive feedback and synergy observed between ML elements and SI learning behaviors suggest that our approach to designing ML-supported SI activities shows potential.

Future research will involve more detailed evaluations of ML4SI with teachers and ML experts. Iterative design improvements will be made based on ongoing feedback collection, with an emphasis on enhancing the user experience and enhancing the integration with educational contexts.

Another essential next step is to analyze the patterns of how teachers and students utilize different ML methods along with SI behaviors. This may involve network analysis, lag sequential analysis, and frequent pattern mining.

We also plan to assess the effects of ML-supported SI learning activities designed by teachers, on students' engagement and development of data/computational literacy. Future work may also analyze the long-term effects of integrating ML into K-12 education, including tracking the development of student's critical thinking and problem-solving skills over time, as well as their readiness for advanced studies or careers in data-driven fields.

ACKNOWLEDGMENTS

We would like to acknowledge our three anonymous study participants, Minh Nguyen for contributing to the system development, and the National Science Foundation (RETTL program award No. 2225227) for supporting this work.

REFERENCES

[1] Babak Alipanahi, Andrew Delong, Matthew T Weirauch, and Brendan J Frey. 2015. Predicting the sequence specificities of DNA-and RNA-binding proteins by deep learning. *Nature biotechnology* 33, 8 (2015), 831–838.

[2] Pierre Baldi, Peter Sadowski, and Daniel Whiteson. 2014. Searching for exotic particles in high-energy physics with deep learning. *Nature communications* 5, 1 (2014), 4308.

[3] Nicholas M Ball and Robert J Brunner. 2010. Data mining and machine learning in astronomy. *International Journal of Modern Physics D* 19, 07 (2010), 1049–1106.

[4] Barbara Rita Barricelli, Fabio Cassano, Daniela Fogli, and Antonio Piccinno. 2019. End-user development, end-user programming and end-user software engineering: A systematic mapping study. *Journal of Systems and Software* 149 (2019), 101–137. https://doi.org/10.1016/j.jss.2018.11.041

[5] Karianne J Bergen, Paul A Johnson, Maarten V de Hoop, and Gregory C Beroza. 2019. Machine learning for data-driven discovery in solid Earth geoscience. *Science* 363, 6433 (2019), eaau0323.

[6] Christopher M Bishop and Nasser M Nasrabadi. 2006. *Pattern recognition and machine learning*. Vol. 4. Springer.

[7] Rita Borgo, Johannes Kehrer, David HS Chung, Eamonn Maguire, Robert S Laramee, Helwig Hauser, Matthew Ward, and Min Chen. 2013. Glyph-based Visualization: Foundations, Design Guidelines, Techniques and Applications.. In *Eurographics (state of the art reports)*. 39–63.

[8] Herman Chernoff. 1973. The use of faces to represent points in k-dimensional space graphically. *Journal of the American statistical Association* 68, 342 (1973), 361–368.

[9] Ton De Jong and Wouter R Van Joolingen. 1998. Scientific discovery learning with computer simulations of conceptual domains. *Review of educational research* 68, 2 (1998), 179–201.

[10] John J. Dudley and Per Ola Kristensson. 2018. A Review of User Interface Design for Interactive Machine Learning. 8, 2 (2018). https://doi.org/10.1145/3185517

[11] Stephen HC DuToit, A Gert W Steyn, and Rolf H Stumpf. 2012. *Graphical exploratory data analysis*. Springer Science & Business Media.

[12] Gintautas Dzemyda, Olga Kurasova, and J Zilinskas. 2013. Multidimensional data visualization. *Methods and applications series: Springer optimization and its applications* 75, 122 (2013), 10–5555.

[13] Steven D Essinger and Gail L Rosen. 2011. An introduction to machine learning for students in secondary education. In *2011 Digital Signal Processing and Signal Processing Education Meeting (DSP/SPE)*. IEEE, 243–248.

[14] Ignacio Evangelista, Germán Blesio, and Emanuel Benatti. 2018. Why are we not teaching machine learning at high school? A proposal. In *2018 World Engineering*

Education Forum-Global Engineering Deans Council (WEEF-GEDC). IEEE, 1–6.

[15] Johannes Fuchs, Fabian Fischer, Florian Mansmann, Enrico Bertini, and Petra Isenberg. 2013. Evaluation of alternative glyph designs for time series data in a small multiple setting. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 3237–3246.

[16] Yolanda Gil, Mark Greaves, James Hendler, and Haym Hirsh. 2014. Amplify scientific discovery with artificial intelligence. *Science* 346, 6206 (2014), 171–172.

[17] Michael Gleicher, Danielle Albers, Rick Walker, Ilir Jusufi, Charles D Hansen, and Jonathan C Roberts. 2011. Visual comparison for information visualization. *Information Visualization* 10, 4 (2011), 289–309.

[18] Thomas RG Green. 1989. Cognitive dimensions of notations. *People and computers V* (1989), 443–460.

[19] Lintang Matahari Hasani, Dana Indra Sensuse, Kautsarina, and Ryan Randy Suryono. 2020. User-Centered Design of e-Learning User Interfaces: A Survey of the Practices. In *2020 3rd International Conference on Computer and Informatics Engineering (IC2IE)*. 1–7. https://doi.org/10.1109/IC2IE50715.2020.9274623

[20] Gwo-Jen Hwang, Haoran Xie, Benjamin W Wah, and Dragan Gašević. 2020. *Vision, challenges, roles and research issues of Artificial Intelligence in Education*. Vol. 1. Elsevier. 100001 pages.

[21] Fethi A Inan and Deborah L Lowther. 2010. Laptops in the K-12 classrooms: Exploring factors impacting instructional use. *Computers & Education* 55, 3 (2010), 937–944.

[22] Alfred Inselberg. 1985. The plane with parallel coordinates. *The visual computer* 1 (1985), 69–91.

[23] Robert JK Jacob and Howard E Egeth. 1976. The face as a data display. *Human Factors* 18, 2 (1976), 189–200.

[24] Michael I Jordan and Tom M Mitchell. 2015. Machine learning: Trends, perspectives, and prospects. *Science* 349, 6245 (2015), 255–260.

[25] Alla Keselman. 2003. Supporting inquiry learning by promoting normative understanding of multivariable causality. *Journal of Research in Science Teaching* 40, 9 (2003), 898–921.

[26] ChanMin Kim, Min Kyu Kim, Chiajung Lee, J Michael Spector, and Karen DeMeester. 2013. Teacher beliefs and technology integration. *Teaching and teacher education* 29 (2013), 76–85.

[27] Stephen Kuhn, Matthew J Cracknell, and Anya M Reading. 2018. Lithologic mapping using Random Forests applied to geophysical and remote-sensing data: A demonstration study from the Eastern Goldfields of Australia. *Geophysics* 83, 4 (2018), B183–B193.

[28] Donald MacGregor and Paul Slovic. 1986. Graphic representation of judgmental information. *Human-Computer Interaction* 2, 3 (1986), 179–200.

[29] Adam V Maltese, Joseph A Harsh, and Dubravka Svetina. 2015. Data visualization literacy: Investigating data interpretation along the novice—expert continuum. *Journal of College Science Teaching* 45, 1 (2015), 84–90.

[30] Lívia S Marques, Christiane Gresse von Wangenheim, and Jean CR Hauck. 2020. Teaching machine learning in school: A systematic mapping of the state of the art. *Informatics in Education* 19, 2 (2020), 283–321.

[31] Ramon Mayor Martins and Christiane Gresse Von Wangenheim. 2023. Findings on teaching machine learning in high school: A ten-year systematic literature review. *Informatics in Education* 22, 3 (2023), 421.

[32] Samuel T McAbee, Ronald S Landis, and Maura I Burke. 2017. Inductive reasoning: The promise of big data. *Human resource Management review* 27, 2 (2017), 277–290.

[33] New York State Education Department. 2023. The New York State K12 Computer Science and Digital Fluency Learning Standards. https://www.nysed.gov/.

[34] Margus Pedaste, Mario Mäeots, Leo A Siiman, Ton De Jong, Siswa AN Van Riesen, Ellen T Kamp, Constantinos C Manoli, Zacharias C Zacharia, and Eleftheria Tsourlidaki. 2015. Phases of inquiry-based learning: Definitions and the inquiry cycle. *Educational research review* 14 (2015), 47–61.

[35] Markus Reichstein, Gustau Camps-Valls, Bjorn Stevens, Martin Jung, Joachim Denzler, Nuno Carvalhais, and fnm Prabhat. 2019. Deep learning and process understanding for data-driven Earth system science. *Nature* 566, 7743 (2019), 195–204.

[36] Alexander Repenning and Andri Ioannidou. 2006. *What Makes End-User Development Tick? 13 Design Guidelines*. Springer Netherlands, Dordrecht, 51–85.

[37] Matthias Rupp, Alexandre Tkatchenko, Klaus-Robert Müller, and O Anatole Von Lilienfeld. 2012. Fast and accurate modeling of molecular atomization energies with machine learning. *Physical review letters* 108, 5 (2012), 058301.

[38] Ismaila Temitayo Sanusi, Solomon Sunday Oyelere, Henriikka Vartiainen, Jarkko Suhonen, and Markku Tukiainen. 2023. A systematic review of teaching and learning machine learning in K-12 education. *Education and Information Technologies* 28, 5 (2023), 5967–5997.

[39] Advait Sarkar. 2016. *Interactive analytical modelling*. Ph. D. Dissertation. University of Cambridge, Cambridge, United Kingdom.

[40] Deborah Seehorn, Stephen Carey, Brian Fuschetto, Irene Lee, Daniel Moix, Dianne O'Grady-Cunniff, Barbara Boucher Owens, Chris Stephenson, and Anita Verno. 2011. *CSTA K–12 Computer Science Standards: Revised 2011*. ACM.

[41] Priti Shah and James Hoeffner. 2002. Review of graph comprehension research: Implications for instruction. *Educational psychology review* 14 (2002), 47–69.

[42] Ngss Lead States. 2013. *Next Generation Science Standards: For States, By States*. The National Academies Press, Washington, DC.

[43] Danny Tang et al. 2019. *Empowering novices to understand and use machine learning with personalized image classification models, intuitive analysis tools, and MIT App Inventor*. Ph. D. Dissertation. Massachusetts Institute of Technology.

[44] Jingwan Tang, Xiaofei Zhou, Xiaoyu Wan, Michael Daley, and Zhen Bai. 2023. ML4STEM Professional Development Program: Enriching K-12 STEM Teaching with Machine Learning. *International Journal of Artificial Intelligence in Education* 33, 1 (2023), 185–224.

[45] Matti Tedre, Tapani Toivonen, Juho Kahila, Henriikka Vartiainen, Teemu Valtonen, Ilkka Jormanainen, and Arnold Pears. 2021. Teaching machine learning in K–12 classroom: Pedagogical and technological trajectories for artificial intelligence education. *IEEE access* 9 (2021), 110558–110572.

[46] Xiaoyu Wan, Xiaofei Zhou, Zaiqiao Ye, Chase K. Mortensen, and Zhen Bai. 2020. SmileyCluster: Supporting Accessible Machine Learning in K-12 Scientific Discovery. In *Proceedings of the Interaction Design and Children Conference*. Association for Computing Machinery, New York, USA, 23–35.

[47] De Wang, Feiping Nie, and Heng Huang. 2014. Unsupervised feature selection via unified trace ratio formulation and k-means clustering (track). In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2014, Nancy, France, September 15-19, 2014. Proceedings, Part III 14*. Springer, 306–321.

[48] Matthew O Ward. 2008. Multivariate data glyphs: Principles and practice. *Handbook of data visualization* (2008), 179–198.

[49] David Weintrop, Elham Beheshti, Michael Horn, Kai Orton, Kemi Jona, Laura Trouille, and Uri Wilensky. 2016. Defining computational thinking for mathematics and science classrooms. *Journal of science education and technology* 25 (2016), 127–147.

[50] Pascal Wilhelm and JJ Beishuizen. 2003. Content effects in self-directed inductive learning. *Learning and instruction* 13, 4 (2003), 381–402.

[51] Ian H Witten and Eibe Frank. 2002. Data mining: practical machine learning tools and techniques with Java implementations. *Acm Sigmod Record* 31, 1 (2002), 76–77.

[52] Pak Chung Wong and R Daniel Bergeron. 1994. 30 years of multidimensional multivariate visualization. *Scientific Visualization* 2 (1994), 3–33.

[53] Zacharias C Zacharia, Constantinos Manoli, Nikoletta Xenofontos, Ton De Jong, Margus Pedaste, Siswa AN van Riesen, Ellen T Kamp, Mario Mäeots, Leo Siiman, and Eleftheria Tsourlidaki. 2015. Identifying potential types of guidance for supporting student inquiry when using virtual and remote labs in science: A literature review. *Educational technology research and development* 63 (2015), 257–302.

[54] Xiaofei Zhou, Jingwan Tang, Michael Daley, Saad Ahmad, and Zhen Bai. 2021. "Now, I want to teach it for real!": Introducing machine learning as a scientific discovery tool for K-12 teachers. In *Artificial Intelligence in Education: 22nd International Conference, AIED 2021, Utrecht, The Netherlands, June 14–18, 2021, Proceedings, Part I 22*. Springer, 486–499.

[55] Xiaofei Zhou, Jingwan Tang, Michael Daley, Saad Ahmad, and Zhen Bai. 2021. "Now, I Want to Teach It for Real!": Introducing Machine Learning as a Scientific Discovery Tool for K-12 Teachers. In *Artificial Intelligence in Education*, McNamara D. Sosnovsky S. Luckin R. Dimitrova V. Roll, I. (Ed.). Springer, Cham, 486–499. https://doi.org/10.1007/978-3-030-78292-4_39

[56] Xiaofei Zhou, Jingwan Tang, Beilei Guo, Hanjia Lyu, and Zhen Bai. 2022. Challenges and Design Opportunities in Data Analysis for ML-Empowered Scientific Inquiry-Insights from a Teacher Professional Development Study. In *Proceedings of the 16th International Conference of the Learning Sciences-ICLS 2022, pp. 847-854.* International Society of the Learning Sciences.

[57] Xiaofei Zhou, Jessica Van Brummelen, and Phoebe Lin. 2020. Designing AI learning experiences for K-12: Emerging works, future opportunities and a design framework. *arXiv preprint arXiv:2009.10228* (2020).