International Journal of Artificial Intelligence in Education
https://doi.org/10.1007/s40593-022-00292-4

ARTICLE

ML4STEM Professional Development Program: Enriching K-12 STEM Teaching with Machine Learning

Jingwan Tang[1] · Xiaofei Zhou[2] · Xiaoyu Wan[1] · Michael Daley[1] · Zhen Bai[2]

Accepted: 17 March 2022

Abstract
The advances of machine learning (ML) in scientific discovery (SD) reveal exciting opportunities to utilize it as a cross-cutting tool for inquiry-based learning in K-12 STEM classrooms. There are, however, limited efforts on providing teachers with sufficient knowledge and skills to integrate ML into teaching. Our study addresses this gap by proposing a professional development (PD) program named ML4STEM. Based on existing research on supporting teacher learning in innovative technology integration, ML4STEM is composed of Teachers-as-Learners and Teachers-as-Designers sessions. It integrates an accessible ML learning platform designed for students with limited math and computing skills. We implemented this PD program and evaluated its effectiveness with 18 K-12 STEM teachers. Findings confirm that ML4STEM successfully develops teachers' understanding of teaching STEM with ML as well as fosters positive attitudes toward applying the ML as an in-class teaching technology. Discussions on the implications of our findings from ML4STEM are provided for future PD researchers and designers.

Keywords ML4STEM · Professional development program · K-12 STEM teaching · Machine learning

✉ Zhen Bai
zbai7@ur.rochester.edu

Jingwan Tang
jtang21@u.rochester.edu

Xiaofei Zhou
xzhou50@ur.rochester.edu

Xiaoyu Wan
xwan3@ur.rochester.edu

Michael Daley
mdaley@warner.rochester.edu

1 Present Address: Warner School of Education, University of Rochester, Rochester, USA

2 Department of Computer Science, University of Rochester, Rochester, USA

Published online: 15 April 2022

Introduction

With the rapid development of artificial intelligence (AI), there is a growing need for preparing 21st-century students with basic AI literacy (Evangelista et al., 2018). Machine learning (ML), as a key branch of AI, makes predictions and uncovers key insights from big data. It has advanced technologies in a wide range of applications such as personal assistants, self-driving cars, and healthcare and recently, the function and role of ML in education is being increasingly addressed (Gil et al., 2014).

There are emerging efforts to introduce AI to K-12 education (Touretzky et al., 2019; Marques et al., 2020), with the main focus on extending CS and Engineering curricula with AI/ML knowledge (e.g., Kahn and Winters (2017), Sabuncuoglu (2020), Sperling and Lickerman (2012), and Druga (2018)). These efforts required dedicated teachers with strong AI/ML content knowledge to be successful and often reach limited numbers of students in a school. Recent efforts started incorporating AI/ML learning experiences with science contexts (e.g., Zhang et al. (2019), Lin et al. (2020), Sakulkueakulsuk et al. (2018), Evangelista et al. (2018), and Zimmermann-Niefield et al. (2019)). These efforts also have limitations and barriers as they could easily overwhelm in-service STEM teachers who already have heavy teaching workloads, and they may also lack a computing background for learning and teaching AI/ML (Marques et al., 2020). Therefore, we propose a new approach to introducing ML to K-12 classrooms, which is to integrate basic and accessible ML technologies with crosscutting discovery tools for a broad range of STEM subjects, before revealing more in-depth ML concepts and methods. One recent study (Zimmermann-Niefield et al., 2019) has shed light on integrating ML with the non-CS STEM curriculum and showed K-12 students' potential for data collection and model evaluation through athletic moves. Yet there remain a research gap in Professional Development (PD) programs to prepare K-12 STEM teachers to adopt ML in their classroom, as highlighted by a recent review study about the lack of suggestions and empirical evidence on "the training of instructors to prepare them adequately for the application of the ML-based instructional materials in the classroom" (Marques et al., 2020).

As a new discovery tool, ML may provide novel learning opportunities that engage students in authentic and evidence-based scientific inquiry that constitutes the core science practices for K-12 STEM standards, such as asking questions, developing models, analyzing and interpreting data, and engaging in argumentations with evidence (States, 2013). Imagine if, with ML-empowered discovery tools accessible to novice learners without much experience in ML, a biology teacher can facilitate students to learn, discover and make sense of various ecological phenomena from online datasets. Unexpected or puzzling patterns may spur students' curiosity, leading students to ask big questions about the puzzling phenomenon and obtain a more profound understanding through intrinsic investigation.

Preparing teachers to integrate ML into STEM classrooms is challenging mainly for three reasons: (1) lack of pedagogical knowledge for teachers to apply ML as a scientific discovery (SD) learning tool in STEM teaching (e.g., Sullivan et al. (2020) and Zhang et al. (2019)); (2) technological barriers for K-12 teachers to sufficiently apply ML technologies in STEM contexts (Mariescu-Istodor and Jormanainen, 2019); and (3) absence of ML-empowered SD lesson plans that demonstrate how K-12 STEM learning can benefit from the introduction of ML. Without such knowledge and resources, teachers might be misplaced with fear or underrate their capabilities to teach with ML, which may impact their belief and attitude to apply ML their instruction plans (e.g., Inan and Lowther (2010), Kim et al. (2013), Mouza (2009), Marques et al. (2020), and Tang (2019)).

To address these challenges, we propose ***ML4STEM***, a novel PD framework that aims to prepare K-12 teachers to integrate ML as a new discovery tool in STEM teaching. ML4STEM is composed of two sessions: Teachers-as-Learners (TaL) and Teachers-as-Designers (TaD). The TaL session aims to develop teachers' initial technical understanding of ML in the K-12 STEM context by going through an example ML-empowered SD learning activity named SmileyDiscovery [1]. SmileyDiscovery (Zhou et al., 2021) is extended based on an existing ML learning environment called SmileyCluster [2], which has been proved to be highly accessible to high school students with limited math and computing backgrounds (Wan et al., 2020). It utilizes novel face-based data visualization technologies to teach *k*-means clustering, a representative unsupervised ML method widely used in STEM domains such as biology, climate, and medicine (Tan et al., 2016). The TaD session aims to facilitate teachers to obtain pedagogical knowledge of utilizing key ML components involved in clustering to fulfill scientific inquiry processes. This is achieved through a co-design workshop that supports the K-12 teachers to collaboratively design authentic ML-empowered STEM lesson plans by applying different ML components from SmileyDiscovery (e.g., multidimensional data exploration, similarity computation, pattern recognition etc.) in various stages of scientific inquiry (e.g., conceptualization, exploration, interpretation etc.). ML4STEM is developed based on the PD literature on fostering teacher learning for "technology integration", including the learning goals and the principles of effective learning activities (Kopcha, 2012; Mouza, 2009; Tondeur et al., 2012). We designed the learning goals by adapting the **TPACK** (Mishra and Koehler, 2006) framework, which is an established PD framework that explicates the essential knowledge components for effective teaching with technology in various K-12 contexts. In addition, we adopted five principles of effective learning activities to mitigate identified technological and pedagogical learning barriers: ***learning by design***, ***hands-on experience with technology***, ***modeling of technology use***, ***collaborative participation***, and ***reflection***.

We applied the ML4STEM PD framework in the *Integrating Technology with STEM teaching* course in a school of education at a research-based university in New York State. The study took place in two 75-minute online classes in two consecutive weeks, with 18 K-12 in-service STEM teachers. We evaluated teachers' knowledge development through the established TPACK framework (Koehler & Mishra, 2009) and belief change. Findings confirmed the effectiveness of ML4STEM in supporting K-12 teachers to obtain key PD skills of (1) teachers' technological understanding of the ML components of SmileyDiscovery, (2) applying *k*-means clustering in different subject matters, (3) pedagogical understanding of strengths and limitations of using ML components offered by SmileyDiscovery for in-class SD learning instructions, and (4) applying ML components in SmileyDiscovery to design inquiry-based learning activities for STEM topics. Teachers' confidence and interest in ML

[1] https://pacific-headland-34136.herokuapp.com/

[2] https://augnitionlab.github.io/FaceOverlay_Publish/

integration progressively increased through the TaL to TaD sessions. Insights on patterns for knowledge development and belief change, design guidelines for future ML-STEM integration, as well as limitations are discussed in this paper.

The main contributions of this study are three-fold: (1) provide a novel PD framework for integration of ML into K-12 STEM teacher learning; (2) offer the corresponding measures that help evaluate teacher learning outcomes of ML knowledge development and ML-STEM integration; and (3) present the design guidelines for future ML-STEM integrations in K-12 PD programs.

Related Work

Our work builds on research in ML, K-12 STEM education, and teacher PD. We first introduce the existing work that connects ML methods with SD practices, a promising approach to integrating ML into the K-12 STEM curriculum. Then, we discuss recent efforts on preparing K-12 teachers to learn and teach ML in class, mainly focusing on the challenges in teacher learning and the limitations of the PD program design. Finally, we present literature that informs our ML4STEM PD program in three sub-sections: (1) TPACK PD framework, (2) pedagogical principles for effective learning activities, and (3) strategies of PD design.

ML-enhanced Scientific Discovery

As an essential branch of AI, ML plays an increasingly important role in scientific discovery for the STEM communities (Langley, 2000) by suggesting novel correlations, accelerating data search processes, and revealing complex patterns from a large amount of data (Gil et al., 2014). There are three main types of ML methods: supervised learning, unsupervised learning, and reinforcement learning.

Unsupervised learning draws an inference from datasets without data labels, which provides a convenient and exploratory lens into some basic ML concepts such as multi-dimension feature space and similarity comparison (Wan et al., 2020). As one of the most popular unsupervised learning algorithms, *k*-means clustering has a wide range of applications in scientific discoveries across different STEM domains, such as hydrology (Ay & Kisi, 2014), ecology (Kupfer et al., 2012), chemistry (Perini, 2013), biology and archaeology (Romesburg, 2004). Researchers have applied *k*-means clustering to identify underlying data patterns (Romesburg, 2004), and conduct scientific discoveries, such as uncovering shared features within each cluster of objects (Essinger & Rosen, 2011; Evangelista et al., 2018), inferring correlations between attributes (Skapa et al., 2012), and supporting feature selection for modeling (Wang et al., 2014).

Teaching ML in K-12 STEM Contexts

Current research on professional learning for teaching ML in K-12 STEM is limited (Marques et al., 2020). We identified three studies pursuing this line of inquiry. One study (Vazhayil et al., 2019) designed a teacher education program for computer science teachers to learn how to introduce ML in their schools. Another study (Zhang et al., 2019) trained teachers to integrate computing thinking with science teaching by using a model named *Logic Programming*. Both projects attempted to develop teachers' understanding of ML through merely direct instructions, such as presenting slides and providing the textbooks; however, it might fail to enable teachers to connect ML applications with their teaching practice. On the contrary, (Sullivan et al., 2020) designed an active learning environment for elementary school teachers to learn the integration of ML into teaching. Specifically, they required teachers to work in dyads and think about the connections between computing practices (e.g., creating algorithms and writing codes) and specific content areas (e.g., English, mathematics, and science). However, this study lacks a systematic description of the PD program design and the associated evaluation approach to examine the process of teacher learning. Such knowledge is essential for guiding future researchers or practitioners in designing an ML integration PD program. Thus, our study aims to address this gap by exploring guidelines and evaluation methods for a PD program that develops K-12 teachers' competence in teaching with ML.

TPACK PD Framework

TPACK, developed by Koehler and Mishra (Koehler & Mishra, 2009), is a well-recognized PD framework that explicates seven knowledge components for technology integration into teaching: ***Content knowledge (CK***) - subject matter to be taught and learned in class (e.g., mathematics, literacy, and history); ***Pedagogical knowledge (PK)*** - practices of teaching knowledge, including classroom management, learning styles & characteristics of students, and instructional methods; ***Technology knowledge (TK)*** - productive operation of technology in theories and in practices; ***Pedagogical content knowledge (PCK)*** - pedagogical knowledge applied to a specific subject area (e.g., using inquiry-based learning approach to teach biodiversity); ***Technological content knowledge (TCK)*** - the content ideas that are enhanced or constrained by the technology; ***Technological pedagogical knowledge (TPK)*** - the pedagogical practices are supported or not supported by a specific technology; ***Technological pedagogical and content knowledge (TPCK)*** - the holistic technology integration that emerges from interactions among content, pedagogy, and technology knowledge. It is considered the basis of good teaching with technology and requires an understanding of the pedagogical strategies that use technology in constructive ways to teach content.

ML4STEM mainly focuses on four particular dimensions, namely TK, TCK, TPK, and TPCK, as they are suggested by prior work to have unique connections with technologies (Mouza, 2009).

Pedagogical Principles for Effective Learning Activities

There is a rich literature on pedagogical principles of effective learning activities that facilitate knowledge development for learners with varied backgrounds. We introduce five principles in this section that inform the design of ML4STEM to address the challenges for integrating ML to STEM classrooms.

Learning by Design

Designing the instructional materials for technology use requires teachers to explore the technology for specific educational purposes, therefore encouraging them to systematically make connections between technology, subject matters, and the means of teaching (Koehler & Mishra, 2005a). It is, therefore, argued to be the most desirable method for developing TPCK (Koehler & Mishra, 2005b; Bakah et al., 2012; Tondeur et al., 2012; Polly et al., 2010). Studies have shown that such activities not only develop teachers' understanding of TPCK (Tondeur et al., 2012; Polly et al., 2010) but also help teachers develop positive attitudes toward implementing technology in practices (Voogt et al., 2011; Cober et al., 2015). ML4STEM adopts the *learning by design* principle by including designing ML-empowered SD lesson plans that utilize basic ML components and SD learning activities in the PD program.

Hands-on Experiences with Technology

Unlike traditional lecture-based PD programs, constructivist researchers view the nature of knowledge as dynamic rather than static, and argue that teachers can gain a meaningful understanding of the technology integration by engaging in the technology-facilitated activities themselves (Darling-Hammond et al., 2017; Goktas et al., 2008). Previous studies demonstrated that hands-on experience provides teachers opportunities to develop understanding of the concepts and skills required for the technology operation (Tearle & Golde, 2008) and builds their technical competencies (Kim et al., 2013; Mouza, 2009).

Modeling of Technology Use

Modeling of technology use refers to the exemplar showing the connections between content, pedagogy, and technology. This is important for teachers, especially at the early stage of developing an understanding of how new technology can be adapted to subjects teaching (Huizinga et al., 2014), as well as teachers' interests in technology adoption (Haydn & Barton, 2007; Tondeur et al., 2012). If teachers have no prior knowledge of technology use, it is impossible for them to construct the meaning and apply it for teaching because such information is not stored in their cognitive structures (Bruner et al., 1966). ML4STEM adopts modeling of technology use by providing teachers tutorial videos that demonstrate examples of using the chosen ML tool to carry out SD learning activities.

Collaborative Participation

Many studies supported the benefit of working in groups when learning about the educational use of technology (Angeli & Valanides, 2009; Darling-Hammond et al., 2017; Kopcha, 2012; McKenney et al., 2015). The involvement of technical experts is significant for teachers to digest the technical concepts or skills and then develop an understanding of using them in practice. Studies showed that working with technical experts can improve teachers' access to technology, provide a clear vision for using technology for instruction, and foster teachers' belief about technology integration during the planning and implementation of a technology-enhanced learning environment (Kopcha, 2012). In addition to technical experts, research found that collaborating with peer teachers brings benefits to using technology for teaching content areas (Kali et al., 2015), reducing anxiety associated with learning (Angeli & Valanides, 2009), and broadening understanding of different teaching approaches with technologies (Darling-Hammond et al., 2017).

Reflection

Reflection is an active process for exploring the potential of technology integration for class teaching, which can bring transformative changes to teacher learning after incorporating it with technology (Boud et al., 1996). Learning activities such as hands-on learning only provide teachers with limited experiences of what the technology can do, but reflection activities have demonstrated to be helpful to grow teachers' in-depth understanding through critical thinking and contemplation of the connections between the use of technology and teachers' own teaching practices (Webster-Wright, 2009; Tondeur et al., 2012; Jang, 2008). ML4STEM adopts several reflection practices including discussing the strengths and limitations of ML technologies in teaching (Tearle & Golde, 2008; Matuk et al., 2015) and writing the journal (Tondeur et al., 2012) to help teachers enhance understanding and make connections.

Strategies of PD Program Design

Besides the design of learning activities, previous research provides valuable suggestions on the overall design of PD programs for improving teachers' learning of teaching with technology. One of the common PD program design strategies is to ***include an extensive duration***. Several studies have shown that professional learning does not occur as a one-time effort (Kim et al., 2013; Mouza, 2009). Thus, it is necessary to provide sufficient time for teachers to experience the technology and to transform such experience into learning. The other common strategy is ***introducing progressiveness in the learning process***. According to Koehler and Mishra (Mishra & Koehler, 2006), different dimensions of knowledge are not developed simultaneously. Usually, TK is developed first, followed by TCK and TPK (no particular order), and finally TPACK. Therefore, in order to achieve the true integration of technology, pedagogy, and content knowledge, incremental supports should be provided to satisfy the progressive needs of change (Kim et al., 2013). ML4STEM adopts both strategies by offering a two-session teacher learning experience, *Teachers-as-Learners* and *Teachers-as-Designers* (Kali et al., 2018),with efficient reflections after each session.

Table 1 Goals of teacher learning in the ML4STEM PD program

Goals	Definitions in ML Integration
TK	Knowledge of ML concepts and methods (*e.g., k-means clustering*) that inform the ML-enhanced learning tool as well as the skills of operating it.
TCK	Knowledge of what subject matters or content activities can be enhanced or constrained by the ML-enhanced learning tool (*e.g., using ML to discover how different environmental factors affect the development of organisms*).
TPK	Knowledge of what teaching and learning strategies can be supported by the ML-enhanced learning tool (*e.g., using ML to support inquiry-based learning*).
TPCK	Knowledge of how to use the ML-enhanced learning tool via pedagogical strategies to instruct student learning in a specific content topic (*e.g., using ML to scaffold inquiry-based learning activities that explore the relationships between environmental factors and the growth of organisms*).
Beliefs	Attitudes toward applying the ML-enhanced learning tool in subject teaching and student learning.

The Design of ML4STEM Professional Development Program

ML4STEM PD framework aims to help teachers understand how to apply ML in K-12 STEM teaching, with two specific learning goals: (1) ***Knowledge development***: Prepare K-12 teachers with sufficient knowledge to utilize ML as a new discovery tool for STEM teaching relevant to the implementation of innovations, and (2) ***Change of belief***: Facilitate teacher's belief change in applying ML into STEM class teaching. For the knowledge development goal, ML4STEM adopts the TPACK framework, specifically focusing on TK, TCK, TPK, and TPCK due to their relatedness with technologies. Table 1 lists the definitions of these constructs in the context of ML integration.

In this section, we will first introduce the ML-empowered SD learning tool adopted in the ML4STEM PD framework, and then describe the detailed design of the ML4STEM PD framework.

ML-empowered Scientific Discovery Tool

As part of the ML4STEM PD program, our study used SmileyDiscovery (Zhou et al., 2021) by adapting an accessible ML learning platform named SmileyCluster for high school students (Wan et al., 2020). Preliminary findings showed that SmileyCluster can effectively introduce basic concepts and methods of *k*-means clustering to high school students with limited math and computing background (Wan et al., 2020). It utilizes visual reasoning of face-based data visualization, which maps multidimensional data features to facial features, and facilitates the interpretation of patterns by arranging and overlaying faces to compare the similarity of data points within and between clusters.

K-means clustering (Steinley, 2006) is a commonly-used unsupervised ML algorithm to partition multidimensional data into k groups based on the similarities between data points. It works in a few steps: (1) k data points are initially selected as

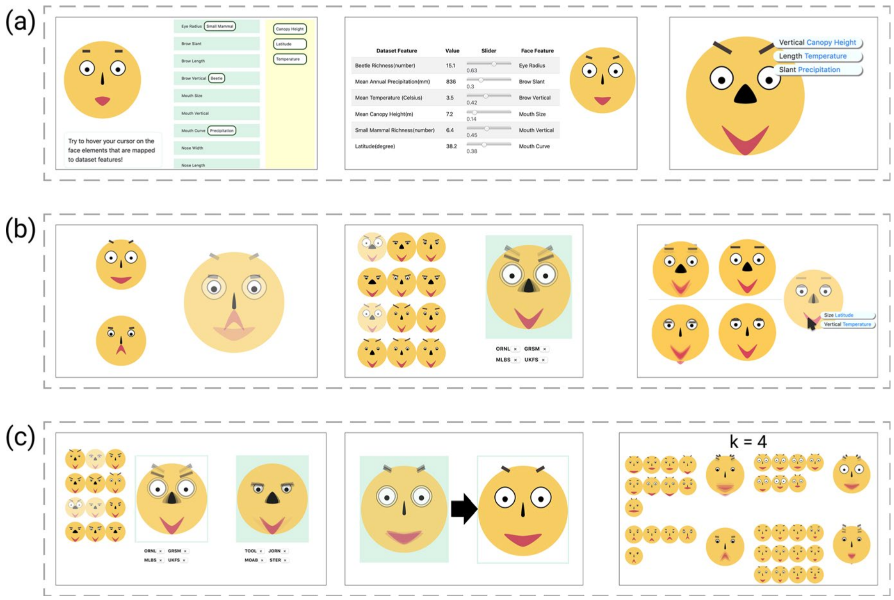

Fig. 1 SmileyDiscovery ML components (from left to right): (a) components to facilitate multidimensional data exploration (modify data-face mapping, explore value mapping, reveal data-face mapping in real-time); (b) components to facilitate pattern interpretation via similarity computation (pairwise comparison, intra-cluster pattern interpretation, inter-cluster pattern interpretation); (c) components to reveal patterns from data (manual clustering, generate centroid, automatic clustering which applies *k*-means clustering to divide data into 4 clusters)

centroids; (2) each data point in the rest of the dataset is assigned to the most similar centroid based on the distance between each data point and centroids; (3) the mean of each cluster is computed as the new centroid to represent the cluster; (4) repeat steps 2 - 3 until the optimal inter-cluster dissimilarities and intra-cluster similarities are achieved. This work focuses on introducing *k*-means clustering as a discovery tool for K-12 STEM teachers to utilize in their teaching. Therefore, the teaching affordances of other areas in ML will not be discussed.

In SmileyDiscovery, three ML components (Fig. 1) of *k*-means clustering are adopted for this study: (1) ***multidimensional data exploration*** (Fig. 1(a)) to facilitate the understanding of the multidimensional problem space; (2) ***similarity computation*** which enables efficient visual comparison between data points to make sense of patterns (Fig. 1(b)) via shared features within a cluster (intra-cluster pattern interpretation), or differentiating features between clusters (inter-cluster pattern interpretation); (3) ***pattern recognition*** (Fig. 1(c)) referring to manual/automatic *k*-means clustering and centroid generation that help learners to carry out investigation for previously raised questions or hypotheses, and may also lead to the conceptualization of new questions and hypotheses.

During a SmileyDiscovery learning activity about ecosystems, for multidimensional data exploration, learners go through all the ecological factors involved in the dataset along with their definitions and generate customized data-face mapping

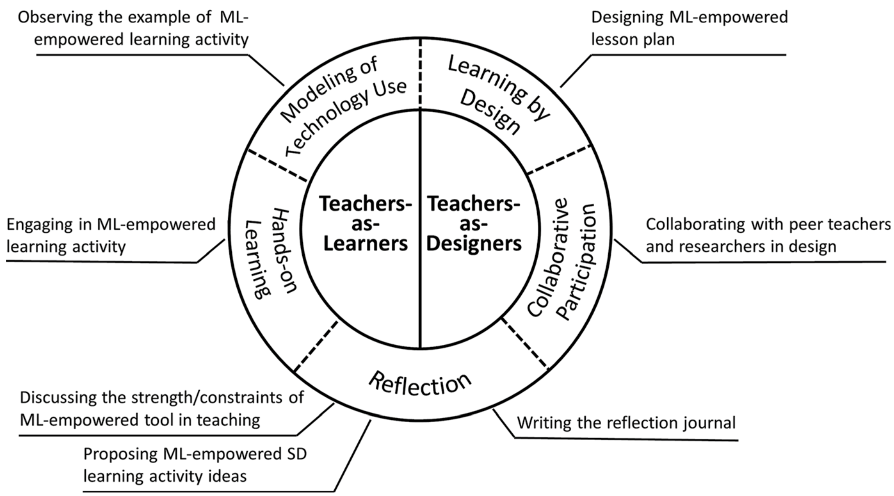

Fig. 2 The ML4STEM PD program includes three components: (1) The inner core indicates the two-session experience that provides progressive scaffolding for efficient teacher learning, named Teachers-as-Learners (TaL) and Teachers-as-Designers (TaD) sessions; (2) The crust is composed of five principles selected based on a literature review to fulfill the intended goals of teacher learning; and (3) The outer layer illustrates corresponding activities implemented to realize each principle

by dragging and dropping data attributes onto facial features. They can also explore the data-face value mapping by interacting with sliders. As to similarity computation, learners move from the pairwise comparison (i.e., overlay and compare two data points of two representative field sites) and the groupwise comparison (i.e., overlay and compare a group of data points of similar field sites), to comparing the cluster centroids. For pattern recognition, learners start with recognizing intra- and inter-cluster patterns from two manually generated clusters of field sites from a subset of data points. After getting more familiar with clusters, centroids, and pattern interpretation, learners will recognize patterns automatically revealed by k-means clustering.

ML4STEM PD Framework

We constructed the ML4STEM PD program based on previous literature on supporting teacher learning of technology integration. See Fig. 2 for the overall structure of the ML4STEM PD framework.

Teachers-as-Learners (TaL) Session

The goal of TaL session is to prepare teachers with an initial understanding of TK, TCK, and TPK as well as to promote their interest in applying ML as an SD tool in teaching. This section includes three main PD principles: modeling of technology use, hands-on learning, and reflection.

Modeling of Technology Use

Modeling of technology use means the exemplar of how ML can be adapted to teaching STEM content. Lacking examples of K-12 learning activities that utilize ML as a discovery tool prevents teachers from understanding TCK and TPK as well as discourages teachers from applying them in classrooms. Thus, providing teachers with specific cases of ML integration is helpful for teachers to see connections between ML and content activities as well as pedagogical practices.

We designed three learning activities informed by this principle for teachers to develop an initial understanding of the application of ML components in STEM teaching. First, teachers ***observe the ML-empowered learning activities*** by watching a tutorial video. We expect teachers to establish a mindset of the integration of ML components with the instructions of scientific inquiry and with the content learning by watching the video. Then, teachers will ***engage in ML-empowered learning activities*** that serve to enhance their understanding of TCK and TPK.

One thing should be noticed, the exemplar of applying ML in STEM teaching should be carefully designed for ensuring a meaningful integration of ML (TK), instructions for supporting SD learning (PK), and STEM content (CK). In our study, we demonstrated the main SmileyDiscovery ML components to a science educator and worked together with the educator to develop three example SD learning activities. The main ML-empowered SD activities include three phases (Table 2).

Hands-on Learning

Hands-on learning refers to learning ML integration in teaching via tactile activities. We designed ***engaging in ML-empowered learning activities***, referring to walking through the pre-designed example activities, which aim to develop teachers' understanding of ML concepts and methods necessary for teaching with ML components (TK). As stated in the Related Work section, hands-on learning is critical for understanding ML technology for teachers with limited computing backgrounds. Further, by trying out the pre-designed ML-empowered activities from a learner's perspective, teachers are able to see what knowledge their students can develop through the same activity (TCK) and how lessons with scientific inquiry can be enhanced by ML methods to effectively support student learning (TPK).

Reflection

Reflection indicates the thinking activities in which teachers connect ML technologies with their own teaching experiences. The goal of incorporating this principle in the TaL session is to help teachers shift their understanding of ML integration from the perspectives of PD program designers to the knowledge situated in their teaching backgrounds. We designed two activities informed by this principle to develop teachers' understanding of TPK and TCK, respectively. One activity asks teachers to ***discuss the strengths and constraints of the ML-enhanced tool in teaching*** in group discussion to critically examine the connections between SmileyDiscovery

Table 2 An example of an ML-empowered SD learning activity

ML components (TK)	Inquiry-based learning (PK)	Learning activities (CK)
Multidimensional data exploration	Orientation and initial conceptualization	Students are introduced to the variables related to the dynamic relationship within ecosystems, such as temperature, annual precipitation, beetle richness, canopy height, etc. They are asked to generate their initial hypothesis based on the observation of two field sites and their prior knowledge.
Pairwise comparison and group-wise comparison	Initial Investigation	By comparing data about several ecosystems, students conduct an initial investigation on how those ecological features influence each other dynamically.
Automatic clustering	Further Investigation and conceptualization	By looking into each cluster, students are expected to develop scientific explanations based on the cumulative evidence from further investigation through pattern recognition and interpretation.

ML components and the means of teaching. The guided questions include “After experiencing the ML-empowered SD learning activity, what are your favorite features and how they might help you in class teaching?”, “What features do you think should be improved for better supporting student scientific discovery and why?”, along with other questions. The other activity requires teachers to ***propose ML-empowered SD learning activity ideas*** which can encourage them to creatively think about the applications of the ML components to their teaching subjects. We expect that through reflection teachers can go beyond the modeling use of the SmileyDiscovery ML components and develop a transformative understanding by connecting these components to their own teaching practices.

Teachers-as-Designers (TaD) Session

The goal of the TaD session is to prepare teachers with an integrative understanding of TPCK as well as develop positive attitudes toward applying ML in teaching. The design of this session composes three principles: Learning by design, collaborative participation, and reflection.

Learning by Design

Learning by design refers to learning of ML integration by designing ML-empowered learning activities. We asked teachers to design ML-empowered SD lesson plans for encouraging them to think about ML components, content topics, and instructions for scientific inquiry in an integrated manner (TPCK).

To support the design process, we created a design canvas (Fig. 3) by using LucidChart, a web-based diagram software allowing for online collaboration. The collaborative design canvas includes two core elements as scaffolding: a lesson plan design areas (Fig. 3 Area 1 & Area 3) informed by the backward design theory (Wiggins et al., 2005) and the established 5E model of scientific instruction (Bybee et al., 2006) for structuring inquiry-based learning activities (Fig. 3 Area 2). The 5E instruction model forms a full learning cycle comprising five cognitive states: Engage (assess students’ prior knowledge and helps students engage in a new concept), Explore (provide students with a common base of activities within which current concepts, processes, and skills are identified and conceptual change is facilitated), Explain (enable students to describe understanding and pose questions about the concepts they are exploring), Elaborate (challenge and extend students’ conceptual understanding and skills) and Evaluate (assess students’ understanding and abilities and provides opportunities for teachers to evaluate student progress towards achieving the educational objectives). When designing the lesson plan, teachers are requested to map the SD learning activities guided by 5E instruction steps and then select corresponding SmileyDiscovery ML components to align with each instruction step. First, teachers will be guided to discuss and fill out the learning objectives and other related information in Area 1 (Fig. 3). Then they can drag and drop cards that are color-coded for different 5E instruction stages from Area 2 to Area 3 where each step in the lesson plan is illustrated. Major system components will be selected

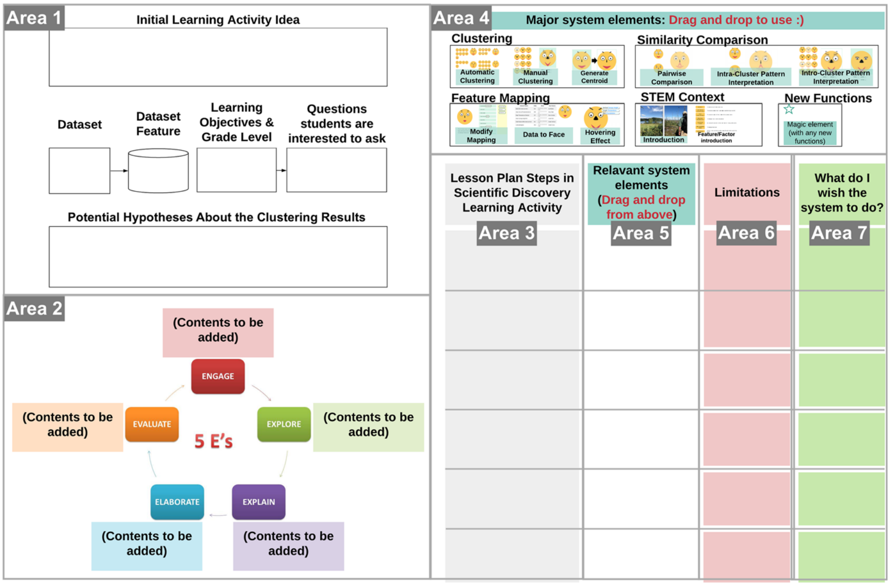

Fig. 3 The design canvas used in the TaD session with (1) Area 1: learning objectives and target learners based on the initial idea and the dataset selected; (2) Area 2&3: the 5E instruction model for teachers to develop each step in the lesson plan; (3) Area 4&5: draggable ML components in SmileyDiscovery for each step in the lesson plan; (4) Area 6&7: limitations of existing SmileyDiscovery components and new designs identified for steps from the lesson plan

from Area 4 and placed on the corresponding row in Area 5 to facilitate different instructional steps. If certain steps can not be fully achieved by the existing system components, teachers can type the limitations (Area 6) and desired features (Area 7).

Collaborative Participation

Collaborative participation refers to the involvement of experts who are responsible for the content, the pedagogy, and the technology in learning activities of using ML in teaching. We integrated this principle with the activity of designing ML-empowered lesson plans in two kinds. First, we allow ***collaboration between peer teachers in design*** by requiring teachers with similar teaching grades and subjects to work in groups. As teachers know well of content and pedagogy knowledge of subject teaching, the interactions between them can provide rich resources for individual teachers to make sense of how to use SmileyDiscovery ML components to support STEM teaching. Second, we encourage teachers to ***collaborate with researchers in design*** by assigning our researchers to each teacher group. We expect the participation of technical experts to support teachers' understanding of ML concepts and of the features in the SmileyDiscovery activity. Moreover, teacher-researcher interaction might produce creative use of the ML-enhanced tool in SD learning activities due to different participants' perspectives.

Reflection

The goal of reflective activities is to encourage teachers to critically examine their attitudes toward applying ML as an SD learning tool in K-12 contexts after participating in the ML4STEM. We ask teachers to ***write reflection journals*** by which they can deeply explore their perceptions of ML technology and its use in STEM teaching. The guided questions to be addressed in journals are: (1) Do you think ML4STEM helps you develop ML knowledge related to utilizing SmileyDiscovery components in STEM teaching? (2) Do you consider SmileyDiscovery as an effective tool for students learning STEM content? (3) How is it possible for you to integrate SmileyDiscovery components in your class?

Methods

We conducted a study to evaluate whether ML4STEM can help teachers develop an understanding of ML integration in STEM teaching, aiming to answer the following two research questions: (1) Can ML4STEM help teachers develop TK, TCK, TPK, TPCK? (2) Can ML4STEM help teachers develop beliefs about ML-empowered STEM teaching?

Research Settings and Participants

We embedded the ML4STEM PD program in a teacher education course -*Integrating Technology with STEM teaching*- which was a component of a Noyce Master Teacher Fellowship Program. Teachers registered for this course for understanding various facets of how technology was and could be integrated into STEM classrooms. Eighteen teachers volunteered to be participants and signed the consent forms prior to the study. There were 7 male and 11 female teachers, between 25 and 54 years old, teaching math (N = 9) and science (N = 9) subjects. Participants teaching grade includes elementary school (N = 3), middle school (N = 7), and high school (N = 8). Regarding the previous experience with AI and ML, 6 participants (33.3%) had little experience while the rest of the participants (66.7%) had no experience.

Study Procedure

We implemented the two-session ML4STEM PD program in two consecutive weeks via the online platform Zoom. Each session lasted for 75 minutes due to the class time constraints.

TaL: Teachers-as-Learners (Week 1)

TaL took place in the first class session which includes the following steps.

1. *Introduction (10-min)*: The researchers introduced the overall procedures of the session.
2. *Observing ML-Empowered Learning Activities (20-min)*: Eighteen teacher participants were divided into four Zoom breakout rooms based on their teaching grade levels. Each group was with one researcher to offer technical support and manage time. Teachers watched a tutorial video that demonstrates the example of one pre-designed ML-empowered SD activity with the wine quality dataset (Cortez et al., 2009).
3. *Engaging in ML-Empowered Learning Activities (30-min)*: Teachers independently completed the pre-designed ML-empowered SD activity with the dynamic ecosystem dataset.
4. *Discussing the Strengths and Constraints of the ML-enhanced tool in teaching (30-min)*: Teachers shared comments on the technology design and brainstormed how the system could be used to support their own ML-empowered SD learning, etc, with the guidance of a researcher. To ensure equal participation, researchers invited all participants to share their thoughts.
5. *Proposing ML-Empowered SD Learning Activity Ideas*: After the class, participants were asked to reflect on their learning by posting three SD learning activity ideas that could be implemented by SmileyDiscovery components on the Blackboard discussion forum.

TaD: Teachers-as-Designers (Week 2)

TaD was conducted in the second class session with the following steps.

1. *Preparation (10-min)*: The researcher played a video tutorial to guide participants in accessing the platform and using the tool for collaboration in groups.
2. *Collaboratively Designing ML-Empowered SD Lesson Plans (50-min)*: Each group discussed to select one SD learning activity design idea from the previously posted ideas. Each group collaboratively created a lesson plan by using the design canvas, with one researcher moderating the discussion.
3. *Discussing the constraints and customization of the ML-enhanced tool in teaching (25-min)*: The researcher guided a group discussion on limitations, potential improvement and customization of the ML components for the designed lesson plans.
4. *In-Class Sharing (25-min)*: Each group shared out the designed lesson plan with all participants.
5. *Writing reflection journals*: After the class, teachers wrote a reflection journal to answer three prompts related to their learning experience about the PD program,

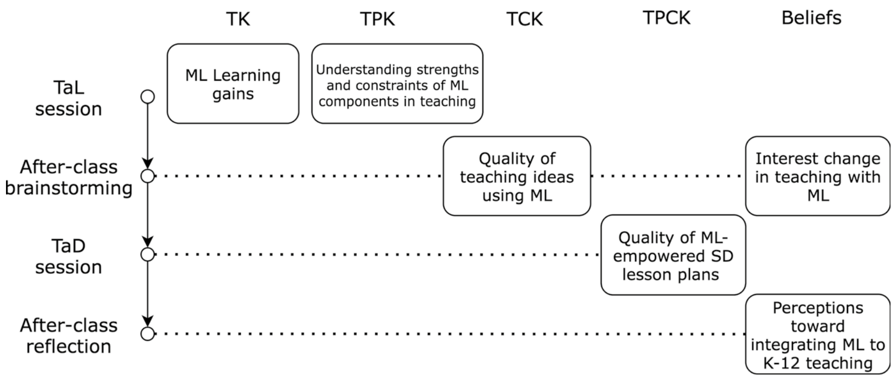

Fig. 4 The ML4STEM measures for different types of knowledge in the TPACK framework

the effectiveness of ML in supporting SD learning, and the feasibility of ML integration into their own class.

Measures & Data Analysis

The measurement methods and data analysis are organized in this section by the order of TaL and TaD sessions. We defined the measure goal of each teacher's learning outcome of ML4STEM by adapting the TPACK framework and SmileyDiscovery ML components. A summary of the measures and the data collection timeline is shown in Fig. 4.

TaL Session

Technology Knowledge (TK) refers to the understanding of basic concepts and methods of *k*-means clustering.

We measured the *learning gain of TK* using pre- and post-study assessments administered prior to and after the TaL session. The assessments contain six open-ended questions about concepts and methods of *k*-means clustering involved in the SmileyDiscovery activity, including multidimensional data exploration, the definition of similarity, similarity comparison, clustering process, centroid, and appropriate k value (Wan et al., 2020). Each question was evaluated on a 0-3 scale using a researcher-developed rubric, reaching a near-perfect inter-rater agreement (Lazar et al., 2017)between two independent raters (Pre-study Cohen's kappa $=0.85$ and post-study Cohen's kappa $=0.83$). We applied z-test to check the data normality and paired t-test for the statistical difference in scores between pre- and post-assessment.

Technology content knowledge (TCK) refers to the understanding of applying SmileyDiscovery ML components to teach disciplinary content knowledge.

To measure the TCK teachers developed after the TaL session, we assessed the *quality of teachers' ML-empowered teaching ideas*. They were collected from teachers' online posts on three teaching ideas that could be supported by SmileyDiscovery ML components. Each posted idea was rated 0-2 scores according to three

Table 3 Codes and definitions for content analysis for teachers' understanding of strengths and limitations of SmileyDiscovery ML components in teaching

***Step1**:Is this turn talking about the strengths/constraints of SmileyDiscovery?*	
Strengths	Teachers reported the benefits or strengths of using SmileyDiscovery ML components in teaching.
Constraints	Teachers reported the constraints of using SmileyDiscovery ML components in teaching.
***Step2**:Which phases of inquiry-based learning did teachers talk about in this turn?*	
Orientation	Teachers talked about instructions related to familiarizing students with the background of the phenomena and the main variables.
Conceptualization	Teachers talked about instructions related to supporting students to propose questions for investigation or generate testable hypotheses.
Investigation	Teachers talked about instructions related to enabling students to analyze data.
Conclusion	Teachers talked about instructions related to drawing conclusions based on investigations.
***Step3**:What ML components of SmileyDiscovery did teachers talk about in this turn?*	

criteria: (1) involving a multidimensional dataset (0.5pt); (2) involving clustering groups (0.5pt); (3) involving appropriate use of ML components for SD activities (1pt).

Two researchers independently coded all the ideas, validated the inter-rater reliability (Cohen's Kappa = 0.76), and computed the average of two ratings as the final score for each idea. We then assessed the quality of TCK for each teacher using a 4-Likert scale: Excellent (all three posts were scored 2pts); Good (two posts were scored 2pts); Fair (only one post was scored 2pts); Poor (none of the posts were scored 2pts).

Technology pedagogical knowledge (TPK) refers to the understanding of applying ML components of SmileyDiscovery to scaffold inquiry-based pedagogy.

We applied content analysis (Lincoln & Guba, 1985) to measure the quality of TPK (i.e. teachers' *understanding of the strengths and limitations of ML in SD teaching*) by coding the focus group interview transcripts in three steps (Table 3).

We first identified related turns on strengths or constraints of using SmileyDiscovery ML components for teaching and then identified its phase aligned with the inquiry-based learning framework (Pedaste et al., 2015), an instruction model for SD teaching. Finally, we open coded SmileyDiscovery ML components mentioned by teachers, with three researchers reaching a consensus on the codes. By combining these three-level codes for each turn, we captured teachers' perceptions of how a particular ML component facilitates or inhibits teaching in SD learning activities.

Beliefs refer to teachers' attitudes toward ML integration in K-12 STEM classrooms regarding its effectiveness in supporting student learning and the feasibility of applying it to teaching. After TaL session, we assessed ***teachers' interest change*** in teaching with ML by analyzing the pre and post self-efficacy survey administered before and after the TaL session. Given the limited duration of TaL and the literature claimed attitude shifting often involves a long episode of learning (Kim et al., 2013), we further measured teachers' perspectives on applying ML in K-12 teaching after the TaD session by conducting thematic analysis of reflection journals.

We assessed teachers' *interest change in teaching with ML* by using the pre and post ML self-efficacy survey with a 7-Likert scale containing five questions adapted from the existing self-efficacy survey for STEM Learning (Glynn et al., 2009; Pintrich & de Groot, 1990). The data is normally distributed based on z-test; therefore a paired t-test was conducted to see if there was a statistically significant increase from pre- to post-assessment.

TaD Session

Technology pedagogical content knowledge (TPCK) refers to the understanding of designing a lesson plan which integrates the SmileyDiscovery ML components, the instructions for SD learning, and a specific STEM content area.

The quality of ML-empowered SD lesson plans was assessed by a rubric adapted from an empirically validated literature that measures TPCK in TCK, TPK, and PCK dimensions (Harris et al., 2010). We first asked researchers of our teams with a solid ML and education background to review each of the designed lesson plans and developed standard versions of (1) The appropriate use of SmileyDiscovery ML components that fulfill the chosen content activities (TCK). For example, one stage of group 4's lesson plan aims to enable students to "explain the features shared by the groups of people with high, medium, and low risk of heart disease (after automatic clustering)" need to be supported by the ML components: intra-cluster pattern interpretation, generated centroid, and inter-cluster pattern interpretation are required. (2) The appropriate use of ML components selected in the lesson plan to support SD activities via the 5E instructional model (TPK) (Table 4); and 3) the alignment of the 5E model and the content activities (PCK) (Bybee, 2009). Second, we rated the quality of the lesson plan on a 4-Likert scale by comparing it with the standard version of each dimension: 1) Excellent (4pts): the lesson plan designed by teachers is fully aligned with the standard version; 2) Good (3pts): Mostly aligned with the standard version; (3) Fair (2pts): partially aligned with the standard version; (4) Poor (1pt): Not aligned with the standard version. Three researchers independently coded four lesson plans and the disagreements were resolved through discussion. The average score of TCK, TPK, and PCK dimensions of an ML-empowered lesson plan represents its quality of TPCK.

To establish a deep understanding of teachers' beliefs about ML integration in K-12 STEM classrooms, we measured ***teachers' perceptions of teaching with ML*** after TaD session. We conducted a thematic analysis of teachers' reflection journals to understand teachers' self-reported effectiveness of ML as a learning tool and the feasibility of integrating ML in class teaching, respectively (Table 5). Two researchers independently coded, and validated the answers to the questions (Cohen's kappa = 1 and 0.92) (Lazar et al., 2017).

Table 4 Standard version of the TPK alignment

Steps	Goals	System components
Engage	Stimulate students interests for engaging in new concepts	**Introduction**: *Introduction, Feature introduction*
Explore	- Get familiar with the concepts to be learned	**Multidimensional data exploration**: *Feature introduction*
	- Make plans for conducting experiments and investigations	**Multidimensional data exploration**: *Modifying mapping relationship, Data to face, Hovering effect*
	- Running experiments	**Similarity computation**:*pairwise comparison, Groupwise comparison*
		Pattern recognition: *Manual clustering, Automatic clustering, Generate centroid*
Explain and Elaborate	Use evidence to explain the relationships between variables	**Pattern interpretation via similarity computation**: *Intra-cluter pattern interpretation, inter-cluster pattern interpretation*
Evaluate	- Evaluate progress or knowledge	Similar to the steps of Explore&Explain&Elaborate
	- Asks additional questions for further understanding of the new concepts	Similar to the steps of Explore&Explain&Elaborate

Table 5 Rubric for content analysis of teachers' self-reported beliefs in reflection journal

Code	Sub-code	Definition
Effectiveness of ML as a learning tool	Effective	Teachers consider ML conducive for STEM learning.
	Less effective	Teachers consider ML having limitations in STEM teaching and require modifications to make it effective.
Feasibility of integrating ML in class teaching	Very likely	Teachers explicitly express their willingness of integrating ML with class teaching, even though they recognize the challenges of teaching with ML.
	Somewhat likely	Teachers are interested in implementing ML in teaching and they want to know more about relevant knowledge before class implementation.
	Somewhat unlikely	Teachers find themselves struggling to propose a teaching topic that fits ML methods, and they are not interested in knowing more about the integration of ML and in-class teaching.
	Very unlikely	Teachers explicitly expressed their unwillingness of using ML in teaching.

Table 6 Paired t-test results for learning gain of ML knowledge (N = 18)

Statement	Pre-survey *M* (*SD*)	Post-survey *M* (*SD*)	*t*	*sig.*
Q1. What does it mean to cluster a dataset?	1.33 (1.36)	2.47 (0.67)	*-2.91*	***.010***
Q2. What is the importance of similarity when clustering a dataset?	0.75 (1.19)	1.50 (1.19)	*-2.03*	*.058*
Q3. What makes two data points similar or dissimilar?	0.31 (0.49)	1.50 (1.22)	*-4.26*	***<.001***
Q4. What is the center point of a group of data points?	0.89 (1.07)	1.75 (1.19)	*-2.67*	***.016***
Q5. Could you order the major steps for the *k*-means clustering algorithm?	0.97 (0.61)	1.56 (0.78)	*-3.58*	***.002***
Q6. Given two different numbers of groups for clustering the same dataset, how do you decide which number of groups gives better results?	0.25 (0.49)	1.14 (1.04)	*-4.05*	***<.001***

Post-hoc Analysis

We further explored the influences of TK (post ML knowledge) on other learning outcomes (TCK and TPK and beliefs) by adopting correlation tests. In addition, we examined if there were teacher background differences (through three teaching grade levels and between two teaching subjects)in the learning outcomes (TK, TCK, TPK, and beliefs) via non-parametric tests.

Results

This section presents the results organized by the intended learning outcomes of the TaL and the TaD sessions accordingly.

TaL Session

Technology Knowledge (TK)

For the *learning gain of TK*, teachers' scores increased in all the TK assessment questions, and the paired t-test results (Table 6) show a significant increase from pre- to post-test for five out of six questions. Teachers achieved a good understanding (above 1.5 points) in post-test on three assessment questions about basic ML concepts of *k*-means clustering (e.g., nature of clustering, centroid).

Questions with relatively low post-test scores (below 1.5 points) are around more advanced ML concepts and methods (e.g., the algorithmic steps for clustering, evaluating the appropriate k value), which were not introduced explicitly but embedded implicitly in the ML-empowered SD learning activities. For example, while exploring the dataset for initial investigation, learners are scaffolded

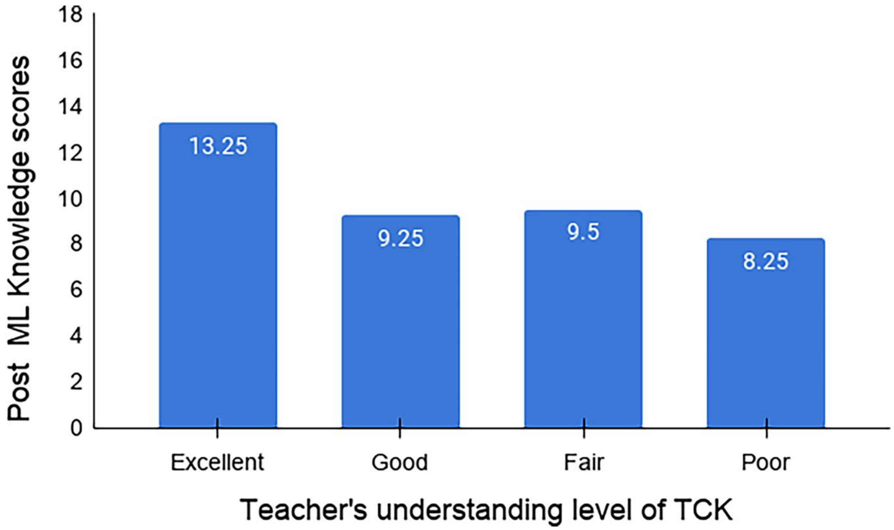

Fig. 5 Descriptive of TCK associated with TK

to manually cluster similar data by prompts following the algorithmic steps for *k*-means clustering. In addition, the learning duration (30-min) for the TaL session is very limited. These might explain why low post-test scores appeared in the more advanced questions.

Technology Content Knowledge (TCK)

To evaluate *the quality of teaching ideas using ML*, 44 ML-empowered SD teaching ideas are collected from 17 participants. After evaluating the quality of ideas, 15 participants reached excellent (N = 2), good (N = 8) or fair (N = 5), and two demonstrated poor understanding of TCK.

We noticed *math teachers were less likely to gain a higher-level understanding of TCK than science teachers*. Seven out of 9 science teachers were above the good understanding, while only 3 out of 8 math teachers achieved that level. Although most math teachers mentioned the advantages of clustering with subject matter (e.g., ”clustering might help 5th graders to learn geometric plane shapes”), they were less likely to explicitly elaborate what features of the dataset will influence clustering results (e.g., specific features that differentiate clusters of rectangles and squares).

In addition, we explored *the relationship between teachers' TK and their TCK*. By examining the mean scores of TK post-test of teachers with different TCK understanding levels (Fig. 5), we found that teachers with an excellent understanding of TCK obtained the highest mean TK score as 13.25, while teachers with poor understanding scored the lowest. However, according to a Spearman correlation test, the association between the post TK and TCK was not statistically significant (p = *.309*).

Table 7 Common themes in strengths and constraints of SmileyDiscovery components for instructing inquiry-based learning

Strengths	Constraints
1. Empowering students to explore new phenomena of their own interest	***1. Insufficient support for meaningful hypothesis generation***
Modify feature mapping: Allow students to select variables and relationships for further investigation	*Modify feature mapping*: Subjective selection of variables might lead to irrelevant results.
2. Supporting students in investigation and interpretation	***2. Cognitive overload in investigation and interpretation***
Face visualization: Enable students to make sense of multidimensional data points	*Face visualization*: Not closely connected to STEM contexts which brings inconvenience in interpreting data in realistic situations.
Manual clustering and automatic clustering: Expedite students to experiment	***3. Hard to discuss with peers if students are using different mapping relationships***
Generating centroids: Help students investigate the differences between clusters and reveal patterns	*Modify feature mapping*: Shape an independent learning environment for conducting investigation which limits students' opportunities of discussing with peers.

Table 8 TPK differences in teacher background

Teacher background	N	Strengths *M* (*SD*)	*Sig.*	Constraints *M* (*SD*)	*Sig.*
(Grade)					
Elementary	3	2.67 (0.58)	*.148*	8.00 (1.00)	***.018***
Middle	7	1.57 (1.13)		5.71 (2.43)	
High	8	1.88 (1.36)		3.25 (1.67)	
(Subject)					
Math	9	1.89 (1.17)	*1.00*	6.44 (2.56)	***.008***
Science	9	1.89 (1.27)		3.33 (1.23)	

Technology Pedagogical Knowledge (TPK)

From the focus group interview in TaL, we identified 122 talk turns showing *teachers' understanding of the strengths and limitations of ML in teaching (TPK)*. Teachers mentioned strengths and constraints of applying SmileyDiscovery ML components to instruct inquiry-based learning for 34 and 88 times respectively. Common themes were generated (Table 7).

A Kruskal-wallis test was performed to explore *relationships between teachers' understanding of TPK and their teaching grade levels* (Table 8). No statistically significant influence of teaching grade was found on the times of reported strengths. Rather, the number of constraints raised per person was significantly differentiated between teaching grades ($H(2) = 8.01$, $p = .018$). Pairwise comparisons using Dunn's test indicated that elementary teachers reported higher numbers of constraints than high school teachers (*p = .023*, adjusted using the Bonferroni correction). No other differences were statistically significant. This suggests that elementary teachers had more concerns about the implementing SmileyDiscovery ML components in teaching. According to the focus group interview, they considered that the system might not be conducive for young-age students who cannot process the large numerical values and facial features representing science variables.

Moreover, a Mann-Whitney test was conducted to understand *how TPK differed between math and science teachers*. Math teachers (*Mdn* =7) mentioned a statistically higher number of constraints than science teachers (*Mdn* =3) (*U = 11.00, p = .008*), while there was no difference in strengths reported (Table 8). This indicates math teachers had more concerns of using SmileyDiscovery ML components for teaching than science teachers.

Beliefs after TaL: Interest Change

Among five questions in pre- and post-test to measure *teachers' interest change of teaching with ML after TaL*, there is a significant increase (*t(17) = -2.40, p = .028*) in "I find machine learning relevant to teaching STEM" (Table 9) indicating teachers' increasing awareness of connections between ML and their teaching subjects. Three items decrease without statistical significance.

The interest in applying ML in teaching was higher for science teachers (M_{pre} =5.11, SD_{pre} =1.13; M_{post} =5.13, SD_{post} =1.11) than math teachers (M_{pre} =4.09, SD_{pre} =1.18; M_{post} =4.33, SD_{post} =1.40) in both pre- and post-test. Regarding question 2 in pre-test, math teachers (*Mdn* =4) were significantly less curious about integrating ML in STEM teaching than science teachers (*Mdn* =5), according to the Mann-Whitney test (*U = 68.50, p = .011*).

Moreover, the interest in applying ML in teaching was higher for middle (M_{pre} =4.77, SD_{pre} =1.00; M_{post} =5.29, SD_{post} =1.30) and high school teachers (M_{pre} =4.95, SD_{pre} =1.12; M_{post} =4.85, SD_{post} =0.88) than elementary teachers (M_{pre} =3.27, SD_{pre} =1.50; M_{post} =3.13, SD_{post} =1.22) in both pre- and post-test. A Kruskal-Wallis test showed a statistical difference between teaching grades for question 3 in post-test (*H(2) =6.38, p = .041*). Dunn's multiple comparison test indicated that elementary teachers scored significantly lower than high school after Bonferroni adjustment (*p = .037*). This meant that elementary teachers were less likely to find ML relevant for their STEM teaching compared to teachers of higher grade levels.

TaD Session

Technology Pedagogical Content Knowledge (TPCK)

We evaluated *the quality of ML-empowered SD lesson plans* collaboratively designed by teachers (Table 11). For a total of 1-4 scale, group 3 obtained the highest score of 3.83, and the other three groups' score were above 3. The high quality indicates four groups demonstrated high TPCK in their lesson plans by integrating ML components, instructions for SD learning, and specific content areas. To provide

Table 9 Paired t-test results for interest change in teaching with ML (N = 18)

Statement	Pre-survey *M* (*SD*)	Post-survey *M* (*SD*)	*t*	*sig.*
Q1. Including machine learning in my teaching is interesting.	4.22 (1.60)	4.67 (1.37)	*-1.64*	*.119*
Q2. I am curious about using machine learning to teach scientific discoveries in my class.	4.61 (1.54)	4.56 (1.69)	*0.16*	*.871*
Q3. I find machine learning relevant to teaching STEM.	4.94 (1.39)	5.50 (1.51)	*-2.40*	***.028***
Q4. I believe that machine learning can be used to make scientific discoveries in my class.	4.56 (1.15)	4.50 (1.65)	*0.16*	*.871*
Q5. It is beneficial to introduce some basic concepts of machine learning in my class.	4.67 (1.28)	4.44 (1.42)	*0.89*	*.386*

a clearer picture of how the lesson plans align the ML components well with the goal of content activities as well as the 5E model, we illustrate the quality of TCK and TPK dimensions respectively.

The ***quality of ML-empowered lesson plan to fulfill content activities (TCK)*** is evaluated based on the TCK rubrics. All four lesson plans exhibit a high level alignment between ”STEM learning content” and the corresponding learning objectives ($M = 3.58$, $SD = 0.36$). There are some common use cases for teachers to deliver teaching content through ML components. With automatic clustering, teachers empower students to efficiently identify major categories based on similarities and representatives of each category. In group 2’s lesson plan, students are required to synthesize the biological characteristics of different species through the ML-empowered bottom-up inductive discovery. Besides, all groups use the representative data points to reduce the cognitive workload for students to investigate further with messy data. For example, group 1 guided students to explore the optimal combination of different categories efficiently by only considering the representative construction materials. Furthermore, group 3 innovatively used different mapping mechanisms between dataset features and facial features for students to observe the dataset and clustering results from different perspectives. By viewing a person’s income along with other different attributes in the clustering results, students can get a very straightforward picture of which factor matters more.

For ***the quality of ML-empowered lesson plans to achieve inquiry-based learning (TPK)***, evaluation results yield that all groups had the average score above 3 ($M = 3.56$, $SD = 0.30$), meaning a relatively high alignment between the chosen ML components and the 5E instructional steps. Specifically, group 3 and group 4, consisting of high school teachers, showed higher alignment scores than the other two groups. The high scoring alignment elements are ”Elaborate” with its matching ML components - pattern interpretation supported by similarity computation features. For example, all four groups supported ”Elaborate” instructional step with intra- or inter-cluster pattern interpretation to help students answer the initial inquiry.

The common misalignment of the components lies between the “Explore” and “Explain” steps with its corresponding ML components, such as comparing the facial feature pairwise or groupwise to identify the similarity/difference with/between groups. Four groups either missed related system components supporting “Explain” the pattern, or misplaced the clustering in “Explain”. For example, group 2 supported the “Explain” step with “manual clustering” and “automatic clustering” components, without further interpretation with “pairwise comparison”, which cannot support their instructional goal of “defining certain groups of the organisms that are similar”. This might be due to teachers’ lack of deeper understanding of how different clustering results should be analyzed differently.

Beliefs after TaD: Perception

To show *teachers’ perceptions of teaching with ML* after the PD program, Table 10 provides an overview of teachers’ beliefs about integrating ML with teaching reported after the TaD session. Thirteen out of 17 teachers valued *ML’s*

Table 10 Teachers' perceptions of ML integration reported in reflection journal

Teachers' self-reported effectiveness of ML in supporting SD learning (N = 17)				
Effective		Less effective		Not respond
13		3		1
Teachers' self-reported feasibility to apply ML in their class (N = 17)				
Very likely	Somewhat likely	Somewhat unlikely	Very unlikely	Not respond
6	6	3	1	1

effectiveness on SD learning for three aspects. First, ML can be used to help students solve scientific problems and learn concepts of different STEM subjects (N = 9). Second, the integration of ML with teaching prepares students for an innovative way of data exploration, analysis and interpretation, which are key skills for students to develop according to the Next Generation Science standards (State, 2016) (N = 10). Third, the ML-empowered learning experience is novel and interesting, which can motivate students to be actively involved in learning activities (N = 3). Two math teachers and one science teacher considered ML as less effective. Among them, the elementary school math teacher reported a great deal of scaffolding and time is required to familiarize students with the ML tool in a constructive manner, which, however, "seemed like a lot of work for very little pay off".

For *the feasibility of applying ML in their own classrooms*, six participants explicitly expressed their willingness to integrate ML with teaching in future (Very likely), even though some of them recognized the challenges for preparing students to navigate with ML technology smoothly. Another six participants were somewhat likely to use ML in their teaching, showing their strong interest in implementing ML in class, but insufficient knowledge hampered them to do so. Nevertheless, this group of teachers were motivated to experience more ML methods and learn more about how to design ML-empowered learning activities. Three teachers were somewhat unlikely to apply ML in teaching. One of them was math teachers, explaining that ML methods were not adaptive to his teaching; while the other two were science teachers, with the concern that it was too complicated for lower-level grade students to understand the transferability of facial features and variables.

Additionally, five teachers expressed that *their attitudes toward learning and using ML for teaching changed after participating in the TaD session*. For example, one participant wrote "for this module, I will be completely honest that I was a bit overwhelmed with the SmileyDiscovery activity at first. However, after this past class (TaD), I will say that creating a scenario and designing the activity really helped me in understanding".

We further explored ***the influences of TK on teachers' perceptions***. We applied the Spearman's correlation and found a significant relationship between the post ML knowledge and the feasibility of applying ML in classrooms, *rs (16)*

Table 11 Four lesson plans designed by teachers collaboratively

Lesson Plan Topic	Dataset	Learning Objectives	Big Questions	ML-Empowered SD Learning Activities
Construction materials and flood resistance	Different physical properties (e.g., absorbency, hardness) of a set of construction materials.	Investigate and compare the different properties of construction materials; Determine what combinations of construction materials will resist floods the best.	1. Which properties are the most important in determining construction materials' flood resistance? 2. What construction materials work together most effectively?	1. Use multidimensional data exploration to raise questions and initiate investigation for how to classify construction materials by observing each dimensions in the dataset; 2. Run automatic clustering to identify typical types of materials based on their physical properties; 3. Compare the centroid from each cluster to decide the optimal combination of construction materials to prevent floods.
Physical characteristics and evolution	Different biological characteristics (e.g., the number of legs) of a set of organisms.	Differentiate organisms based on given biological characteristics.	1. Why do some organisms have different numbers of legs? 2. How do scientists use biological characteristics to differentiate different species?	1. Explore the organism dataset by comparing the similarities between an invasive tick and other arachnids to decide if they belongs to the same species; 2. Experiment with manual clustering to contemplate what biological features define certain groups as similar among the organisms being studied; 3. Run automatic clustering to identify what features are unique to each cluster by intra- and inter-cluster pattern interpretation. 4. Identify the cluster that matches the invasive tick through each organism cluster's unique features.

Table 11 (continued)

Lesson Plan Topic	Dataset	Learning Objectives	Big Questions	ML-Empowered SD Learning Activities
Influential factors of income	Different attributes about people including their income, age, education, etc.	Identify patterns in people's income based on their other specific attributes.	1. What effect do these attributes (e.g., education, occupation, gender, race) have on a person's future income?	1. Select the income attribute and two other attributes that a person doesn't have control over (e.g., race) and map them onto facial features; 2. After generating patterns on the entire dataset by automatic clustering, identify similarities and differences between clusters; 3. Select the income and two other attributes that a person has more control over (e.g., education) and map them onto facial features; 4. After a new round of automatic clustering, interpret the clustering results; 5. Elaborate on the differences between two rounds of clustering and think about what features have larger impact and the corresponding actions.

Table 11 (continued)

Lesson Plan Topic	Dataset	Learning Objectives	Big Questions	ML-Empowered SD Learning Activities
Risks factors for heart disease	Different healthcare attributes about people including their age, smoking history, BMI, etc.	Identify risk factors for heart disease and how people can protect themselves against heart disease.	1. Who in my family is at most risk for heart disease? 2. How should I decrease my risk for heart disease?	1. Introduce a list of people's healthcare attributes to students. Select the ones that they think are the most influential to heart diseases and map them to facial features; 2. Identify some potential risk factors by observing and manually clustering a few faces into groups with high, medium and low level risks of heart disease; 3. After automatic clustering on the entire dataset, students compare centroids from clusters with high, medium and low level of heart disease risks; 5. Update hypothesis on the most influential risk factors for heart diseases.

$= 0.84$, $p < .001$. This suggests the better teachers develop TK, the stronger willingness of teachers to integrate ML in class teaching.

Discussion

The results confirmed the overall effectiveness of the ML4STEM PD program in supporting key aspects of learning goals for integrating ML into K-12 STEM classrooms, including TK, TCK, TPK, and TPCK. ML4STEM PD program promotes belief change, represented by the positive development for both self-efficacy of understanding ML concepts and methods (TaL session) and perspective of integrating ML in teaching (TaD session).

Emerging themes connect back to our design principles: (1) Values of learning by design and collaborative participation; (2) Effectiveness of hands-on learning, (3) Necessity of extended learning time; (4) Recommendations for modeling of technology use; (5) Teacher background differences.

Values of Learning by Design and Collaborative Participation

Similar to the previous study of technology integration (Koehler & Mishra, 2005a), designing the ML -empowered SD learning activities plays a vital role in helping teachers develop an integrated understanding of TPCK. A quote from the reflection journal of a middle school math teacher is representative: "*Having the opportunity to design the lesson plan from the bottom-up was very influential in my understanding.*"

Also, our study supports the positive effects of collaborative participation in learning activity design, echoed with the previous research revealing that teachers' anxiety of using the new educational tool could be reduced through designing learning activities with peers (Angeli & Valanides, 2009; Voogt et al., 2015). One participant mentioned that "*When we engaged in the co-design workshop, I was nervous that we were going to have a hard time finding a topic that fits into the program, but the fact that five different groups designed a learning activity at varying grade levels and a variety of concepts, it showed me that there is potential for many ML-empowered learning activities to be created.*" Another participant pointed out in the reflection journal: "*I liked the collaborative aspect of having teachers with the same backgrounds together. It is easy for us to communicate with and understand each other and have a common goal*". These suggest the importance of grouping teachers with similar grade levels or subjects together to form a professional learner community when learning ML integration in class.

Effectiveness of Hands-on Learning

Compared to other traditional educational technologies, gaining knowledge about the integration of ML-enhanced learning tools is more challenging, especially for teachers without computing backgrounds, due to the complexity and abstractness of ML concepts. Our study demonstrates the effectiveness of hands-on learning activities on supporting novice teacher learners to develop a moderate understanding of ML technologies (TK), evident by significant learning gains from the pre to post ML knowledge tests. This is also supported by teachers' feedback in reflection journals, saying that interacting with SmileyDiscovery activity allowed them to make sense of *k*-means clustering. Hands-on learning opportunities through interacting with ML-empowered SD activities helped lower the barrier of teachers' TK development (Tearle & Golde, 2008).

Necessity of Extended Learning Time

We also found that the development of teachers' TK remains at a moderate level. Nearly half of teachers reported in reflection journals that they could not go beyond basic understanding to know the algorithm, and they were unclear about how the computer works for the clustering or the mathematical abstraction of the algorithm through the system. Teacher's belief also showed that insufficient time in TaL for teachers to fully digest ML concepts and relate them to their in-class teaching, as a teacher reported in reflection journal after the TaD session that "ML is one of those areas where there would have to be great deals of scaffolding and time in order to have students be able to use it in a constructive manner." Although learning the algorithm is not the goal of the PD program, the continuous demand of expanding teachers' technological knowledge (TK) needs to be supported in the future PD program design. This is because TK plays a critical role in teacher learning in ML contexts, influencing the development of an integrated understanding of TPCK, and impacting teachers' belief about applying ML into their own class, as discovered in the post-hoc analysis. One potential reason that impedes teachers from digesting the TK could be the limited time (30-min) for ML-empowered SD learning experience in TaL session. As reported by reflection journals, teachers felt overwhelmed by the information in the first session, and they needed "*really dive into what SmileyDiscovery ML components are capable of and implement the system to target at student level.*" While recognizing the importance of extensive duration for PD (Kim et al., 2013; Mouza, 2009), we only created two sessions for the ML4STEM, 75 minutes for each, due to class time constraints and course planning. Future PD program designers should offer adequate learning duration for teachers to gain an in-depth understanding of ML. Also, more hands-on informed activities (e.g., embodied interaction design for teachers' ML learning (Opel et al., 2019) and advanced ML components (e.g., automatic clustering, inter-cluster pattern interpretation) can be provided to support teacher learning.

Recommendations for Modeling of Technology Use

The principle, modeling of technology use, provides teachers opportunities to understand TPCK related to ML integration at an earlier stage. This is evident by the integrative mechanisms of TPCK within the ML-empowered SD lesson plans designed by teachers are similar to the example activities provided for them. Given teachers lacking the mental structure (Bruner et al., 1966) of such knowledge, experiencing an existing exemplar helps them for schema acquisition (Bruner et al., 1966) which is important for further sense-making activities. To better support teacher learning, we suggest two recommendations of applying this principle for future ML4STEM PD program design.

First, provide diverse examples of ML-empowered learning activity to match teachers' subjects and grade levels. This is because teachers who find the exemplar activities resonating with their teaching practices can better understand using ML in teaching. For example, a science teacher wrote in the reflection journal: "*The activity that I enjoyed most was the Ecology and Climate module. With my background in Biology, I was able to relate a lot of the variables and vocabulary that were used*". In comparison, our study's math teachers encountered more challenges in developing an integrated understanding of TPCK, given that no math contents were involved in the three pre-designed learning activities. Similarly, the hesitation of applying ML components to elementary school might be due to the same reason. Said by an elementary teacher: "*I think I would need an actual example of how it (SmileyDiscovery) might be used with primary (school) students*." Given the preliminary stage of the study, we only included pre-designed activities on selected STEM topics. While findings show that most STEM teachers were able to assimilate the pre-designed activities' structures and transfer them to different content areas, a wider range of customized exemplars may be needed to better support teachers with diverse teaching backgrounds (e.g., math).

Second, provide diverse examples of the mechanisms integrating ML as SD tools in STEM teaching. Our study presented teachers with three pre-designed learning activities which, in effect, share the structure of the TPCK integration. Because of that, teachers tended to follow it and were "not comfortable breaking away from the certain flow quite yet." This resulted in the structure of four ML-empowered lesson plans, created by teachers, were consistent with the pre-designed learning examples. As a teacher mentioned in the reflection journal: "*The creativity of designing the lesson plan can be stifled by the exemplar activities.*" According to the theory of improvisation, establishing and expanding repertoire of TPCK (i.e., patterns of TPCK archived in the long-term memory) is one of the essential cognitive processes necessary for stimulating improvisational acts (Biasutti, 2017). Thus, we suggest that future PD program designers offer teachers various structures of TPCK to inspire teachers' creativity.

Teacher Background Differences

Teaching with ML in Mathematics

Compared with science teachers, K-12 math teachers possessed more concerns about applying ML as a teaching tool (TPK), had a lower understanding of integrating ML with content activities (TCK), and showed lower interest in implementing ML in their classes (Beliefs). One reason for such differences is the lack of considerations of math contents within the ML4STEM PD program design. As stated above, teachers' initial understanding of the TPCK is largely influenced by the exemplar learning activities provided for them. Since no modelling of the applications in mathematics was offered, math teachers found it difficult developing knowledge related to using ML to support students learning mathematics. As a result, they were less likely than science teachers to consider ML as an effective learning tool to be implemented in K-12 contexts.

Another potential reason might be the differences in teachers' conceptions of mathematics and science. The focus group and reflection journals reveal an opinion that "the SmileyDiscovery ML components seem not to adapt to math contents." Informed by such ML methods as similarity computation and pattern recognition, SmileyDiscovery ML components are conducive for supporting activities like generating hypotheses, conducting investigations, analyzing, and interpreting data. Such practices, in effect, are aligned perfectly with K-12 science education standards (State, 2016). While we are not suggesting these two disciplines are distinguished, they contain some differences in school education from teachers' perspectives. A previous study reported that math teachers highly embraced rationalism while science teachers ranked empiricism first when discussing the values of teaching subjects the in K-12 school curriculum (Bishop et al., 2006). Such a difference in the conceptions of subjects teaching, in effect, can influence their willingness to engage in educational innovations (Andersen & Krogh, 2010). If teachers find the new ways of teaching do not match with their subject-specific flavors, they would be reluctant to use them. This is not to say that ML as a discovery tool does not fit K-12 mathematics in general. Since few studies have explored this integration, future efforts can be made through the cooperation between K-12 math teachers and computer science researchers.

Teaching with ML in Elementary Schools

In contrast with middle and high school teachers, elementary teachers were more critical about using ML as a teaching tool (TPK) and less interested in applying ML for their class teaching (Beliefs). Nevertheless, they acknowledged the strengths of ML methods in supporting student learning while showing more concerns about the interactive design of the ML-enhanced learning tool. According to the focus group interview, elementary teachers considered that symbolic mapping and relational reasoning (i.e., using facial features as analogs of variables) can confuse young students who are still at Piagetian's concrete thinking stage (Cantu & Herron, 1978). This concern, however, is based on teachers' prior teaching experience and assumptions of their students' cognitive levels, which can be mitigated when seeing the increased learning performance of students brought by the implementation of new educational innovations. According to (Guskey, 2002), teachers' beliefs might not occur with the PD *per se* but occur after successfully implementing new practices in classrooms. A prior study showed that symbolic mapping and relational reasoning were cognitive processes enhanced by children's early literacy and mathematics (Collins &

Laski, 2019), which indicates the future research direction to implement ML-empowered SD learning for young kids.

Limitations

This study contains several limitations. ***First***, our study held a small sample size (N = 18), which constrains statistical power and generalizability of the findings to make inferences for a broader context (Button et al., 2013).Therefore, we have to be careful when interpreting the findings, and the preliminary results are informative for future studies involving larger sample sizes with longer PD sessions. ***Second***, the exposure for teachers to learn the knowledge of teaching with ML is short. On the one hand, teachers can not fully understand TPACK; on the other hand, it provides fewer insights on the developmental trajectory of teachers' learning experiences. Thus, this requires future researchers to extend the implementation duration of ML4STEM PD program, providing teachers with more chances to engage in the learning activities actively. ***Third***, the measures of learning outcomes in this study are inconsistent across two sessions. For example, we assessed TCK and TPK at the individual level for the first session while at the group level for the second session. Also, when analyzing teachers' beliefs, we applied pre and post tests for teachers' interest change after the first session while used thematic analysis of reflection journals for teachers' perceptions after the second session. ***Fourth***, this study only employed one ML-enhanced learning environment to scaffold the ML4STEM PD framework, lacking other ML platforms as supplementary to provide more solid evidence for confirming the generalizability of the ML4STEM PD framework. Future researchers might consider implementing it to support the learning of ML integration by using alternative platforms, such as KNIME, which enables SD activities with a broad range of ML methods (Berthold et al., 2009).

Conclusion

To empower K-12 teachers to utilize ML advances in their STEM teaching, we propose ML4STEM, a PD program grounded in TPACK framework (Mishra & Koehler, 2006) for teachers' effective technology integration. Major design principles utilized include learning by design, collaborative participation, and hands-on learning. An evaluation study with 18 K-12 STEM teachers confirmed the effectiveness of ML4STEM to progressively develop teachers' knowledge and interest in applying ML as a pedagogical tool for content teaching. Also, we found that middle & high school teachers and science teachers encountered less constraints and developed higher interest than elementary and math teachers. In the end, a list of design implications of PD in ML integration is summarized.

Acknowledgements We would like to thank our collaborators doctoral student Zenon Borys from the Warner School of Education, undergrad students Sufian Mushtaq, Saad Ahmad, Abdul Moid Munawar from the Department of Computer Science for helping with system and dataset preparation, Zheng Zhang for helping with data collection, and K-12 teachers for choosing to participate in our study.

Availability of data and material Data is not available to anyone outside of the Institutional Review Board protocol due to ethical restrictions.

Code Availability The code is available on request from the authors.

Declarations

Ethics approval The work presented in this paper was approved by University of Rochester's Research Subjects Review Board (RSRB) under approvals STUDY00003947.

Consent to participate All participants volunteered for participation and signed the consent form prior to the study.

References

Andersen, H. M., & Krogh, L. B. (2010). Science and mathematics teachers' core teaching conceptions and their implications for engaging in cross-curricular innovations. *Nordic Studies in Science Education*, *6*(1), 61–79.

Angeli, C., & Valanides, N. (2009). Epistemological and methodological issues for the conceptualization, development, and assessment of ict–tpck: Advances in technological pedagogical content knowledge (tpck). *Computers & Education*, *52*(1), 154–168.

Ay, M., & Kisi, O. (2014). Modelling of chemical oxygen demand by using anns, anfis and k-means clustering techniques. *J. Hydrol.*, *511*, 279–289.

Bakah, M. A., Voogt, J. M., & Pieters, J. M. (2012). Updating polytechnic teachers' knowledge and skills through teacher design teams in ghana. *Professional Development in Education*, *38*(1), 7–24.

Berthold, M. R., Cebron, N., Dill, F., Gabriel, T. R., Kötter, T., Meinl, T., & Wiswedel, B. (2009). Knime-the konstanz information miner: version 2.0 and beyond. *ACM SIGKDD Explorations Newsletter*, *11*(1), 26–31.

Biasutti, M. (2017). Teaching improvisation through processes: Applications in music education and implications for general education. *Front. Psychol.*, *8*, 911.

Bishop, A., Clarke, B., Corrigan, D., & Gunstone, D. (2006). Values in mathematics and science education: Researchers' and teachers' views on the similarities and differences. *For the Learning of Mathematics*, *26*(1), 7–11.

Boud, D., Keogh, R., & Walker, D. (1996). Promoting reflection in learning: A model. *Boundaries of Adult Learning*, *1*, 32–56.

Bruner, J. S., Olver, R. R., & Greenfield, P. M. (1966). *Studies in Cognitive Growth: A Collaboration at the Center for Cognitive Studies*. New York: Wiley.

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376.

Bybee, R. W. (2009). The BSCS 5E Instructional Model and 21st Century Skills. BSCS, Colorado Springs CO.

Bybee, R. W., Taylor, J. A., Gardner, A., Van Scotter, P., Powell, J. C., Westbrook, A., & Landes, N. (2006). The BSCS 5E instructional model: Origins and effectiveness. BSCS 5, 88-98, Colorado Springs, Co.

Cantu, L. L., & Herron, J. D. (1978). Concrete and formal piagetian stages and science concept attainment. *J. Res. Sci. Teach.*, *15*(2), 135–143.

Cober, R., Tan, E., Slotta, J., So, H. J., & Könings, K. D. (2015). Teachers as participatory designers: Two case studies with technology-enhanced learning environments. *Instr. Sci.*, *43*(2), 203–228.

Collins, M. A., & Laski, E. V. (2019). Digging deeper: Shared deep structures of early literacy and mathematics involve symbolic mapping and relational reasoning. *Early Childhood Research Quarterly*, *46*, 201–212.

Cortez, P., Cerdeira, A., Almeida, F., Matos, T., & Reis, J. (2009). Modeling wine preferences by data mining from physicochemical properties. *Decis. Support. Syst.*, *47*(4), 547–553.

Darling-Hammond, L., Hyler, M. E., Gardner, M., & Espinoza, D. (2017). Effective Teacher Professional Development. Palo Alto CA: Learning Policy Institute.

Druga, S. (2018). Growing up with ai: Cognimates: from coding to teaching machines. Ph.D. thesis Massachusetts Institute of Technology.

Essinger, S. D., & Rosen, G. L. (2011). An introduction to machine learning for students in secondary education. In *Digital Signal Processing and Signal Processing Education Meeting, pp. 243–248*.

Evangelista, I., Blesio, G., & Benatti, E. (2018). Why are we not teaching machine learning at high school? a proposal. In *Proceedings of the World Engineering Education Forum – Global Engineering Deans Council (WEEF-GEDC), pp. 1–6. IEEE.*

Gil, Y., Greaves, M., Hendler, J., & Hirsh, H. (2014). Amplify scientific discovery with artificial intelligence. *Science*, *346*(6206), 171–172.

Glynn, S. M., Taasoobshirazi, G., & Brickman, P. (2009). Science motivation questionnaire: Construct validation with nonscience majors. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, *46*(2), 127–146.

Goktas, Y., Yildirim, Z., & Yildirim, S. (2008). A review of ict related courses in pre-service teacher education programs. *Asia Pacific Education Review*, *9*(2), 168–179.

Guskey, T. R. (2002). Professional development and teacher change. *Teachers and Teaching*, *8*(3), 381–391.

Harris, J., Grandgenett, N., & Hofer, M. (2010). Testing a tpack-based technology integration assessment rubric. In *Society for Information Technology and Teacher Education International Conference, pp. 3833–3840. Association for the Advancement of Computing in Education (AACE).*

Haydn, T. A., & Barton, R. (2007). Common needs and different agendas: how trainee teachers make progress in their ability to use ict in subject teaching. some lessons from the uk. *Computers & Education*, *49*, 1018–1036.

Huizinga, T., Handelzalts, A., Nieveen, N., & Voogt, J. (2014). Teacher involvement in curriculum design: Need for support to enhance teachers' design expertise. *J. Curric. Stud.*, *46*(1), 33–57.

Inan, F. A., & Lowther, D. L. (2010). Laptops in the k-12 classrooms: exploring factors impacting instructional use. *Computers & Education*, *55*(3), 937–944.

Jang, S. J. (2008). The effects of integrating technology, observation and writing into a teacher education method course. *Computers & Education*, *50*(3), 853–865.

Kahn, K., & Winters, N. (2017). Child-friendly programming interfaces to ai cloud services. In *European Conference on Technology Enhanced Learning, pp. 566–570. Springer*.

Kali, Y., Levy, K. S., Levin-Peled, R., & Tal, T. (2018). Supporting outdoor inquiry learning (soil): Teachers as designers of mobile-assisted seamless learning. *Br. J. Educ. Technol.*, *49*(6), 1145–1161.

Kali, Y., McKenney, S., & Sagy, O. (2015). Teachers as designers of technology enhanced learning. *Instr. Sci.*, *43*(2), 173–179.

Kim, C., Kim, M. K., Lee, C., Spector, J. M., & DeMeester, K. (2013). Teacher beliefs and technology integration. *Teach. Teach. Educ.*, *29*, 76–85.

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

Koehler, M. J., & Mishra, P. (2005). Teachers learning technology by design. *J. Comput. Teach. Educ.*, *21*(3), 94–102.

Koehler, M. J., & Mishra, P. (2005). What happens when teachers design educational technology? the development of technological pedagogical content knowledge. *J. Educ. Comput. Res.*, *32*(2), 131–152.

Kopcha, T. J. (2012). Teachers' perceptions of the barriers to technology integration and practices with technology under situated professional development. *Computers & Education*, *59*(4), 1109–1121.

Kupfer, J. A., Gao, P., & Guo, D. (2012). Regionalization of forest pattern metrics for the continental United States using contiguity constrained clustering and partitioning. *Ecological Informatics*, *9*, 11–18.

Langley, P. (2000). The computational support of scientific discovery. *International Journal of Human-Computer Studies*, *53*(3), 393–410.

Lazar, J., Feng, J. H., & Hochheiser, H. (2017). Research Methods in Human-computer Interaction Morgan Kaufmann.

Lin, P., Van Brummelen, J., Lukin, G., Williams, R., & Breazeal, C. (2020). Zhorai: Designing a conversational agent for children to explore machine learning concepts. In *Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 13,381–13,388*.

Lincoln, Y. S., & Guba, E. G. (1985). Naturalistic Inquiry Sage.

Mariescu-Istodor, R., & Jormanainen, I. (2019). Machine learning for high school students. In *Proceedings of the 19th Koli Calling International Conference on Computing Education Research, pp. 1–9*.

Marques, L. S., von Wangenheim, G., C., H., & C., J. (2020). Teaching machine learning in school: A systematic mapping of the state of the art. *Informatics in Education*, *19*(2), 283–321.

Matuk, C. F., Linn, M. C., & Eylon, B. S. (2015). Technology to support teachers using evidence from student work to customize technology-enhanced inquiry units. *Instr. Sci.*, *43*(2), 229–257.

McKenney, S., Kali, Y., Markauskaite, L., & Voogt, J. (2015). Teacher design knowledge for technology enhanced learning: an ecological framework for investigating assets and needs. *Instr. Sci.*, *43*(2), 181–202.

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teach. Coll. Rec.*, *108* (6), 1017–1054.

Mouza, C. (2009). Does research-based professional development make a difference? a longitudinal investigation of teacher learning in technology integration. *Teach. Coll. Rec.*, *111*(5), 1195–1241.

Opel, S., Schlichtig, M., & Schulte, C. (2019). Developing teaching materials on artificial intelligence by using a simulation game (work in progress). In *Proceedings of the Workshop in Primary and Secondary Computing Education, vol. 14, pp. 1–2*.

Pedaste, M., Mäeots, M., Siiman, L. A., De Jong, T., Van Riesen, S. A., Kamp, E. T., & Tsourlidaki, E. (2015). Phases of inquiry-based learning: Definitions and the inquiry cycle. *Educational Research Review*, *14*, 47–61.

Perini, F. (2013). High-dimensional, unsupervised cell clustering for computationally efficient engine simulations with detailed combustion chemistry. *Fuel*, *106*, 344–356.

Pintrich, P. R., & de Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, *82*(1), 33–40.

Polly, D., Mims, C., Shepherd, C. E., & Inan, F. (2010). Evidence of impact: Transforming teacher education with preparing tomorrow's teachers to teach with technology (pt3) grants. *Teach. Teach. Educ.*, *26*, 863–870.

Romesburg, C. (2004). Cluster analysis for researchers. Lulu com.

Sabuncuoglu, A. (2020). Designing one year curriculum to teach artificial intelligence for middle school. In *Proceedings of the ACM Conference on Innovation and Technology in Computer Science Education, pp. 96–102*.

Sakulkueakulsuk, B., Witoon, S., Ngarmkajornwiwat, P., Pataranutaporn, P., Surareungchai, W., Pataranutaporn, P., & Subsoontorn, P. (2018). Kids making ai: Integrating machine learning, gamification, and social context in stem education. In *2018 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE), pp. 1005–1010. IEEE.*

Skapa, J., Dvorsky, M., Michalek, L., Sebesta, R., & Blaha, P. (2012). K-mean clustering and correlation analysis in recognition of weather impact on radio signal. In *2012 35th International Conference on Telecommunications and Signal Processing (TSP), pp. 316–319. IEEE.*

Sperling, A., & Lickerman, D. (2012). Integrating ai and machine learning in software engineering course for high school students. In *Proceedings of the 17th ACM Annual Conference on Innovation and Technology in Computer Science Education, pp. 244–249*.

State, N. Y. (2016). Science learning standards http://www.nysed.gov/curriculum-instruction/science-learning-standards. Accessed 15 Jan 2021.

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

Steinley, D. (2006). K-means clustering: a half-century synthesis. *British Journal of Mathematical and Statistical Psychology*, *59*(1), 1–34.

Sullivan, F., Suárez, E., Pektas, E., & Duan, L. (2020). Developing pedagogical practices that support disciplinary practices when integrating computer science into elementary school curriculum. In *Proceedings of the 14th International Conference on the Learning Sciences, vol. 4, pp. 2289–2292*.

Tan, P. N., Steinbach, M., & Kumar, V. (2016). Introduction to data mining. Pearson Education India.

Tang, D. (2019). Empowering novices to understand and use machine learning with personalized image classification models, intuitive analysis tools, and mit app inventor. Ph.D. thesis Massachusetts Institute of Technology.

Tearle, P., & Golde, G. (2008). The use of ict in the teaching and learning of physical education in compulsory education: how do we prepare the workforce of the future?. *Eur. J. Teach. Educ.*, *31*(1), 55–72.

Tondeur, J., Van Braak, J., Sang, G., Voogt, J., Fisser, P., & Ottenbreit-Leftwich, A. (2012). Preparing pre-service teachers to integrate technology in education: A synthesis of qualitative evidence. *Computers & Education*, *59*(1), 134–144.

Touretzky, D., Gardner-McCune, C., Martin, F., & Seehorn, D. (2019). Envisioning ai for k-12: What should every child know about ai?. In *Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, pp. 9795–9799*.

Vazhayil, A., Shetty, R., Bhavani, R. R., & Akshay, N. (2019). Focusing on teacher education to introduce ai in schools: Perspectives and illustrative findings. In *International Conference on Technology for Education (T4E), pp. 71–77. IEEE*.

Voogt, J., Laferriere, T., Breuleux, A., Itow, R. C., Hickey, D. T., & McKenney, S. (2015). Collaborative design as a form of professional development. *Instr. Sci.*, *43*(2), 259–282.

Voogt, J., Westbroek, H., Handelzalts, A., Walraven, A., McKenney, S., Pieters, J., & De Vries, B. (2011). Teacher learning in collaborative curriculum design. *Teach. Teach. Educ.*, *27*(8), 1235–1244.

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

Wang, D., Nie, F., & Huang, H. (2014). Unsupervised feature selection via unified trace ratio formulation and k-means clustering (track). In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases, pp. 306–321. Springer*.

Webster-Wright, A. (2009). Reframing professional development through understanding authentic professional learning. *Rev. Educ. Res.*, *79*(2), 702–739.

Wiggins, G., Wiggins, G. P., & McTighe, J. (2005). Understanding by Design Ascd.

Zhang, Y., Wang, J., Bolduc, F., Murray, W. G., & Staffen, W. (2019). A preliminary report of integrating science and computing teaching using logic programming. In *Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, pp. 9737–9744*.

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. In *International Conference on Artificial Intelligence in Education, pp. 486–499. Springer*.

Zimmermann-Niefield, A., Turner, M., Murphy, B., Kane, S. K., & Shapiro, R. B. (2019). Youth learning machine learning through building models of athletic moves. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children, pp. 121–132*.