Hybrid Models for Knowledge Tracing: A Systematic Literature Review

Andrea Zanellati, Daniele Di Mitri, Maurizio Gabbrielli, and Olivia Levrini

***Abstract*—Knowledge tracing is a well-known problem in AI for education, consisting of monitoring how the knowledge state of students changes during the learning process and accurately predicting their performance in future exercises. In recent years, many advances have been made thanks to various machine learning and deep learning techniques. Despite their satisfactory performances, they have some pitfalls, e.g., modeling one skill at a time, ignoring the relationships between different skills, or inconsistency for the predictions, i.e., sudden spikes and falls across time steps. For this reason, hybrid machine-learning techniques have also been explored. With this systematic literature review, we aim to illustrate the state of the art in this field. Specifically, we want to identify the potential and the frontiers in integrating prior knowledge sources in the traditional machine learning pipeline as a supplement to the normally considered data. We applied a qualitative analysis to distill a taxonomy with the following three dimensions: knowledge source, knowledge representation, and knowledge integration. Exploiting this taxonomy, we also conducted a quantitative analysis to detect the most common approaches.**

***Index Terms*—Educational data mining, hybrid models, informed machine learning (IML), knowledge tracing (KT), systematic review.**

I. Introduction

LEARNER modeling—also called student modeling—is a widely studied problem due to its relevance in various technologies to enhance learning, including intelligent tutoring systems (ITS) [1] and adaptive educational hypermedia systems (AEHS) [2]. The problem's relevance has its roots in the theories for individualized learning, studied since the 1980s by Cohen et al. [3] and Bloom [4], which prove its effectiveness compared to traditional classroom learning. This motivated the search for technological support and strategies for learner modeling, which could promote individualized learning.

A learner model (LM) is an abstract representation of the learner that considers cognitive and noncognitive characteristics. According to Vagale and Niedrite [5], the LM includes all information available for the system on the user and keeps active user accounts within the system, i.e., it keeps both static and dynamic information about the learner. Specifically, static information is data that are independent from the student' interactions with the system, e.g., personal data or pedagogical and preference data collected once and which stay unchanged during the system utilization. On the other hand, dynamic information is data collected while the students are interacting with the system and they are involved in their learning process. It can refer to the student's performance, i.e., student achievements during the course session, and the actual state of the student's knowledge concepts and skills. The dynamic data component in the LM determines a continuous flow of collecting and updating data about the learner.

As for the methodological approaches to tackle the learner modeling problem, there are two main families [6]. The first set of approaches relies on psychometric methods, e.g., item response theory [7] and cognitive diagnostic models [8], which are mostly based on static data. However, in the last decades, technological advances opened the possibility of collecting dynamic data while the student interacts with a learning system. This attracted the interest of computer scientists in knowledge tracing (KT) [9], which can be described as tracking how students' knowledge states change while they are learning and forecasting their performance in future exercises.

However, there are two main challenges connected to the KT problem. The first one concerns its complexity due to its interdisciplinarity nature. According to Abyaa et al. [10], learner modeling is challenging since it is based on intertwining education science, psychology, and information technology. This led them to suggest that an ideal learner model can be built through the following three steps: the identification and selection of the learner's attributes that impact on their learning; then, considering the learner's psychological states during their learning process, and finally choose the most suitable technologies, which enable to accurately model each selected attribute. This challenge is inherited by KT as a subclass of learner modeling.

The second challenge is spread out by the limitations that emerge from the implementation of ML techniques in KT. Although purely data-driven techniques for KT achieve satisfactory performance, they have some pitfalls regarding their applicability, reliability, or interpretability that we do not find in psychometric models [11]. They mainly differ from purely data-driven approaches because they are grounded in a theoretical framework. In item response theory, for instance, each item is associated with an a priori difficulty coefficient. Moreover, there is the assumption that learning does not occur during testing. Both these assumptions are used in designing the model.

Manuscript received 15 December 2022; revised 16 May 2023 and 7 November 2023; accepted 14 December 2023. Date of publication 1 January 2024; date of current version 31 January 2024. *(Corresponding author: Andrea Zanellati.)*

Andrea Zanellati and Maurizio Gabbrielli are with the Department of Computer Science and Engineering, University of Bologna, 40126 Bologna, Italy (e-mail: andrea.zanellati2@unibo.it; maurizio.gabbrielli@unibo.it).

Daniele Di Mitri is with DIPF Insitute, 60323 Frankfurt, Germany (e-mail: d.dimitri@dipf.de).

Olivia Levrini is with Department of Physics and Astronomy, University of Bologna, 40126 Bologna, Italy (e-mail: olivia.levrini2@unibo.it).

Digital Object Identifier 10.1109/TLT.2023.3348690

This theoretical advantage in psychometric models led us to suggest a possible methodological framework for addressing the pitfalls emerging by purely data-driven techniques, referring to informed machine learning (IML), introduced by von Rueden et al. [12]. In a nutshell, IML aims to overcome a purely data-driven approach favoring hybrid ML models, which integrate alternative knowledge sources to data in the ML pipeline.

In this article, we want to contribute to the two challenges just presented, by conducting a systematic literature review. Specifically, we aim to highlight if and how the prior knowledge due to the interdisciplinary and complex nature of KT can be used to overcome the limitations, which emerge by using standard ML methods. For this purpose, we want to verify if the paradigm proposed with the IML is effective and productive for KT, i.e., whether it can be applied fruitfully to develop KT models. Therefore, we want to find references in the literature that explicitly consider forms of prior knowledge injection to address the KT problem. On the one hand, we can outline the state of the art in deploying hybrid ML approaches for KT. On the other hand, we aim to point out the current gaps in the literature and suggest new avenues for research

The rest of this article is organized as follows. In Section II, we introduce the background. We describe the main ML techniques used for KT with their critical issues. Then, we outline the main features of the IML paradigm. Section III describes the methodology. We display and motivate the RQs tackled through our systematic literature review. We describe both the literature survey procedure and the classification process. In Section IV, we present the results of our analysis. First, we describe the taxonomy we distilled from the surveying of the papers. It is an adaptation of the one proposed by von Rueden [12] due to our focus on the educational field. Second, we show the classification results of the selected papers gained by applying our IML adapted taxonomy. Section V discusses key insights into the results concerning the RQs. Finally, Section VI concludes this article.

II. Background and Challenges

A. Knowledge Tracing

Formally, the KT problem can be described as follows. Let us consider a learner's history exercise sequence $X = \{(q_1, r_1); (q_2, r_2); \ldots; (q_{t-1}, r_{t-1})\}$, where $\{q_i\}$ is the id for the question answered by the learner at the ith-time step, and $r_i = 1$ if the student provided the correct response to the question q_i, 0 otherwise. The goal of KT is to predict the probability of correctly answering the question q_t at time step t, i.e., computing $P_t(r_t = 1|q_t, X)$. Hence, given the unknown function $f : \mathcal{X} \rightarrow \{0, 1\}$, which associate each learner's history exercise sequence (X, q_t) to 1 (q_t correctly answered) or 0 (otherwise), the KT goal is to determine a function $g : \mathcal{X} \rightarrow [0, 1]$, which is a good approximation of f. The prediction is based on hidden variables, whose values are updated at each time step and which model the student's knowledge state.

There are three main classes of ML techniques widely exploited in the literature and well described by Minn [11]: hidden Markov models, factor analysis models, and deep learning-based models. The main exponent of the first class is Bayesian knowledge tracing (BKT) [9]. In this model, the learner's knowledge state is represented through a set of binary variables for each skill or knowledge component (KC), which assumes the true value if the student is in the learned state. The observed data is the student performance, the latent variables are the student knowledge state for each skill. The truth value of the latent variable corresponding to the skill or KC k depends on four factors:

1) the initial learning factor $p(L_0^k)$, which is the prior probability that a student already masters k;
2) the acquisition factor $P(T^k)$, which is the probability for the student to pass from the unlearned state to the learned state after the next practice opportunity;
3) the guess factor $P(G^k)$, i.e., the probability the student guessed the correct answer despite being in the unlearned state;
4) the slip factor $P(S^k)$, which models the probability that the student makes a mistake despite being in the learned state.

The estimate of student mastery of k, i.e., the student knowledge state for k, is continually updated every time a student responds to an item [9]. In a nutshell, the student knowledge state for k after the nth action of the student, indicated by $P(L_n^k)$, is computed considering both the posterior probability that the student was already in the learning state given the evidence (whether or not the nth action is correct), and the probability that he will make the transition to the learned state if it is not already there. Then, the current student's knowledge state for k is exploited to compute the probability to perform a correct action taking into account the mitigation effect of the slip factor $P(S^k)$ and the positive effect of the guess factor $P(G^k)$.

Although BKT has been used successfully in many systems, it has some limitations, well summarized by Tato and Nkambou [13]. First, BKT relies on Bayesian networks (BNs) [14], for which often the student interactions with relevant concepts need manual labeling, and a priori probabilities have to be defined by domain experts. Moreover, BKT handles binary response data to track changes in the student knowledge state, forcing the system to deal only with exercises that can be easily modeled in a Boolean way. Furthermore, BKT is designed to model one skill or KC at a time, ignoring the interactions between skills and KCs and affecting a single performance.

As for the second class, it worths mentioning performance factor analysis (PFA) [15], which is a logistic regression model to predict accuracy considering the student's number of prior failures and successes on that skill. It is an extension of learning factor analysis [16], designed to model multiple skills simultaneously, i.e., the prediction of the student performance relies on the conjunction or compensation of the skills needed in the performance by summing their contributions. PFA is competitive and outperforms BKT models [11]. However, PFA does not consider important behavioral factors, such as the order of answers and the probability of students guessing or slipping. This may affect the reliability of the models prediction.

The third class of techniques is deep learning-based models, which have recently been widely used for KT, as in many other domains. There are two main approaches in this class:

deep knowledge tracing (DKT) models [17], which is based on recurrent neural networks; and dynamic key-value memory networks (DKVMN) [18], which is a memory-augmented neural network based on two memory matrices to exploit the relationships between underlying concepts and directly output a student's mastery level of each concept. As for the disadvantages of DKT, Yeung and Yeung [19] highlight two main points. First, the model fails to reconstruct the observed input, i.e., the model predicts a failure for a student in a certain skill, despite the observation that a student on the same skill in the input data is a success, and vice versa. Second, the predicted performance for skills across time steps is inconsistent, i.e., there are sudden spikes and falls across time steps. Intuitively, this is both unfavorable and illogical, since students' knowledge state is likely to change gradually over time, rather than fluctuating between mastered and not-yet-mastered. Moreover, neural networks have a high computational cost and are prone to overfitting. Sun et al. [20] pointed out some limitations also for DKVMN models. They ignore both the students' behavior features collected by their interaction with the learning system and the student's learning abilities, which affect the students' knowledge state.

In his review, Minn [11] compares item response theory (IRT), BKT, PFA, and DKT on three dynamic public datasets (ASSSITments 2009–2010 and ASSISTments 2014–2015, derived from the homonymous learning system, and Algebra 2005–2006, released in KDD Cup 2010 competition). He obtained the best performance with the IRT psychometric model, followed by the DKT model. The fact that the IRT model performs better than ML models is surprising in the first instance. However, an explanation for this outcome can be find in a main design difference between IRT and other student models, among which also DKT: IRT explicitly relies on the item difficulty factor. This result supports our idea that there is contextual learning information, distinct from the data directly collected on the student and their performance that can provide insights for enhancing expected predictions.

B. Informed Machine Learning

In Section I, we introduce two challenges. First, the problem of KT lies at the intersection of several disciplines, including pedagogy, psychology, cognitive science, and information technology. Second, as supported by the previous section, standard ML models for KT show performance pitfalls that we do not find in psychometric models, which integrate a theory-ladenness.

We can expand the first issue by affirming that learning cannot be described only with information gathered directly from the learner or about the learner. Learning is influenced by the context in which it takes place, understood as a physical, relational, emotional, and disciplinary space [21]. The relevance of the context on learning has been considered since the first research on ITSs, which are based on domain models, pedagogical models, and tutor–learner interface models, together with the LM [22]. However, these components are usually modeled independently, i.e., as 4 separate parts in the system. Little attention is paid to modeling how one can influence the others. Simplifying with an example, on the one hand, the domain model can be seen as an organizational model of the repository for the system's educational resources. On the other hand, it can be understood as an epistemological model of an area of knowledge that may affect how the student learns, hence affecting also the LM [23].

As for the second point, several contributions in the literature affirm the need to overcome purely data-driven approaches in machine learning [24], [25], especially in those contexts where the phenomenon is very complex, it is difficult to obtain sufficiently large and representative datasets. A priori or a posteriori forms of knowledge, acquired over years of research, are available [26]. All these factors exist for KT: its complexity has been motivated in the previous point; the challenge of quantity and quality of data [27] is quite common in the form of class imbalance [28] (e.g., correct answers on skills difficult to master), and a priori and a posteriori knowledge are usually available and are already used in ITSs and AEHSs. In these cases, it may be worthy to test hybrid learning techniques [29], which can be recognized as a strategy of IML.

von Rueden et al. [12, p.616] defined IML as "learning from a hybrid information source that consists of data and prior knowledge. The prior knowledge comes from an independent source, is given by formal representations, and is explicitly integrated into the machine learning pipeline." Here, the term knowledge is assumed in a computer science perspective, meaning verified information regarding the connections between entities within specific contexts. Moreover, they introduced a taxonomy for IML, outlining a scheme consisting of three types of knowledge sources, eight possible knowledge representations, and four forms of integration, as shown in Table I. However, their paper did not refer to educational case studies. Hence, whether their taxonomy fits with the specificity of KT remains to be explored. We suggest referring directly to their paper for a full description of the terms they introduced in the taxonomy. In Section IV, we display the terms that we have distilled for our taxonomy, which is an adaptation of their proposal as a result of our analysis.

III. Methodology

A. Research Questions

To sum up, we are assuming that the complex nature of KT can be addressed by explicitly taking into account the information sources due to the different disciplines that deal with learning and the situation in which it takes place. This means finding a way to integrate these forms of prior knowledge in data-driven machine learning models. Therefore, we took as a reference the taxonomy proposed by von Rueden et al. [12] for IML, trying to apply it to the specificity of our topic. As already mentioned, in their framework, the authors introduce the following three dimensions: knowledge source, knowledge representation, and knowledge integration (see Table I). They also associate each dimension with an analysis question. Here, we assume them as our research questions, focusing the field of study on the KT problem.

RQ1 Which source of knowledge can be integrated into machine learning models for KT?

RQ2 How is the knowledge represented in those models?

TABLE I
IML TAXONOMY INTRODUCED IN [12]

Source	Representation	Integration
Scientific knowledge	Algebraic equations	Training data
World knowledge	Differential equations	Hypothesis set
Expert knowledge	Simulation results	Learning algorithms
	Spatial invariances	Final hypothesis
	Logic rules	
	Knowledge graphs	
	Probabilistic relations	
	Human feedback	

RQ3 Where is the knowledge integrated into the machine learning pipeline?

We opted for a systematic literature review to highlight, which avenues have already been explored, which trends are more common to design hybrid models for KT, and to identify new research methodological trajectories.

B. Literature Surveying Procedure

To perform our systematic literature review, we followed the PRISMA statement [30]. We included four main databases, which contain relevant literature in the field: ACM Digital Library, IEEE Xplore, Scopus, and Web of Science. They are authoritative databases for the research sector in learning analytics and artificial intelligence in education, for which it is possible to carry out searches with articulated queries and by restricting the search field to some parts of the paper (e.g., abstracts and titles).

The query used to retrieve results in these databases is the following: *("skill development"* OR *"skill acquisition"* OR *"skill assessment"* OR *"knowledge tracking"* OR *"knowledge tracing"* OR *"knowledge assessment")* AND*("machine learning"* OR *"artificial intelligence"* OR*computing* OR *"deep learning"* OR *"learning analytics"* OR *"data mining")* AND*(education* OR*educational)*. The search was limited to the titles, abstracts, and keywords of the documents in the databases to select only papers with the main focus on the topic of our interest. The query consists of the conjunction of the following three parts: the first is for keywords about the learning object under study; the second aims to bind the research methodology reference to machine learning and other related fields; the last one is used to disambiguate the terms knowledge and learning, collocating them into the educational sciences. There is a fourth aspect characterizing our research questions regarding the use and integration of sources of prior knowledge. However, it is not easy to identify related keywords, which are sufficiently general for an automatic filtering process taking this aspect into account. In a sense, one of this research's objectives is identifying, which sources are most used as prior knowledge and which lexicon is used to refer to them. Therefore, the focus on prior knowledge was not considered in the first phase of the PRISMA checklist and was integrated later, as we will describe.

The query was run on August 5, 2022, collecting 1267 documents. Fig. 1 shows the main steps of the systematic review process according to the PRISMA flow. In the top-left box of the diagram, we summarized the numbers of retrieved documents with the query search, divided among the selected databases. After removing duplicates and documents in the form of full books or conference proceedings, the list of potential candidate papers was reduced to 957.

On this set of papers, we carried out a manual screening of titles and abstracts to assess the relevance of our study. Specifically, we considered the following inclusion criteria:

1) the paper has a methodological focus on KT, i.e., aims to describe a technique, an algorithm, or a method to deal with KT problems rather than presenting a digital tech application (serious games, virtual reality systems, web platforms, etc.);
2) the main methodological approach refers to the field of machine learning, computational intelligence, or data science;
3) the data used to build the LM are collected from the student's interaction with learning management system (LMS);
4) the paper refers to human learning.

The first two criteria were chosen to pursue the methodological focus of the RQs. The third and fourth criteria were chosen to explicit a sufficiently broad but defined application target for KT. Specifically, the third criterion narrows the interest of this study to school, academic, and training contexts that use LMS as a teaching support tool. The fourth criterion disambiguates the word "learning," which can be used in AI concerning machine or robot learning. In addition, we considered four exclusion criteria:

1) the full text of the paper is not available in English;
2) the paper presents preliminary results, i.e., it is a position papers or the authors declare that they are describing an exploratory study;
3) the paper has as its main objective a literature review;
4) there are later more updated or complete versions of the paper by the same authors and on the same research project.

The application of these criteria led to the selection of 221 papers considered eligible. This set of papers has gone through a new screening phase on the full text, aimed at selecting only documents with a focus on using and integrating prior knowledge sources in ML tools. Specifically, it was decided to consider the following inclusion criteria:

1) the authors explicitly consider the need to integrate a priori forms of knowledge with methods traditionally used to deal with KT;
2) the paper includes a clear description of the methods, i.e., which prior knowledge is taken into consideration and how this is integrated into the ML pipeline.

To check the first criterion, we seek evidence in the text, particularly in the introduction and conclusion sections, where

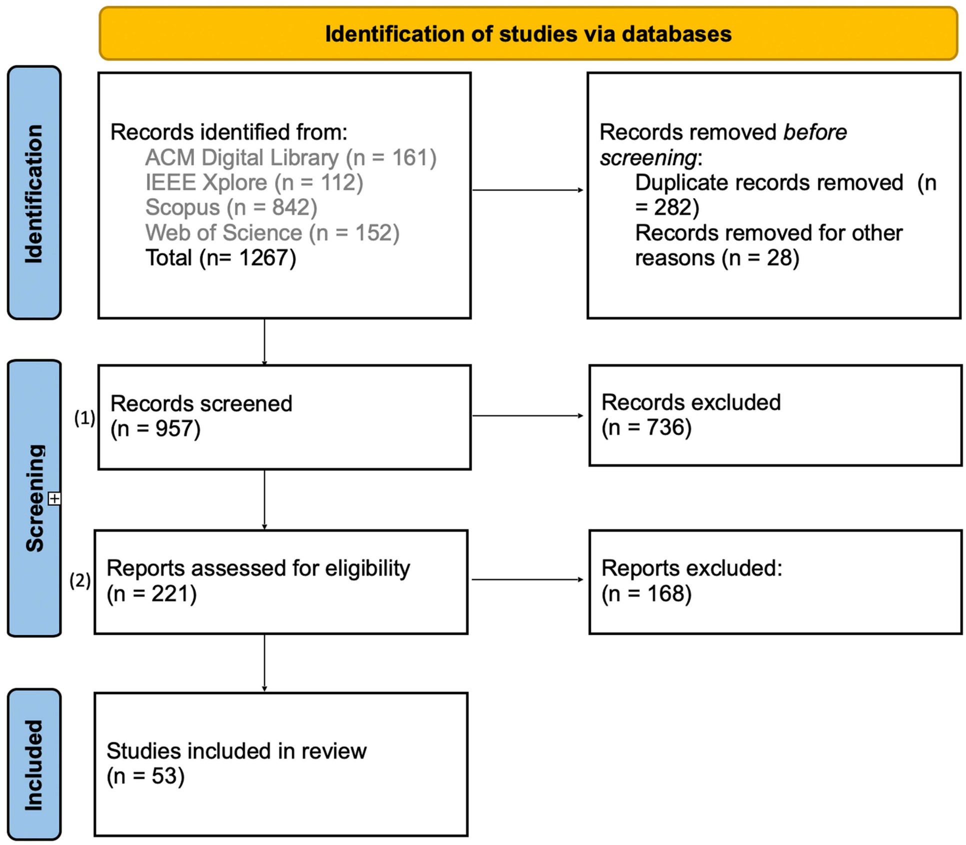

Fig. 1. PRISMA 2020 flow diagram for the screening process. After the identification of the potential candidates, there are two screening steps. The first one 1) consists of applying 4 inclusion and 4 exclusion criteria considering only papers' titles and abstracts. The second 2) has been conducted with two inclusion criteria to focus on the prior knowledge injection problem, considering the papers' full texts. The selection criteria are described in Section III.

authors usually state their main contributions. For the second criterion, the methodology section of each paper was examined.

As a final result, we identified the 53 papers included in this review. All selected papers present case studies in which a close match occurs between the "informed" models and authoritative datasets exploited for KT (e.g., ASSISTment, KDD cup 2010, or data collected with commonly used LMS). The experiments described in the selected papers perform comparably to or better than the purely data-driven models, taken as a reference benchmark. This motivates our interest in exploring potentials and gaps of prior knowledge source integration into the ML pipeline.

C. *Classification Process*

To classify the 53 papers considered eligible for the review, we tested and refined the IML taxonomy by von Ruedend et al. [12], which we have already introduced in Section II.

We felt comfortable using their notation about knowledge integration, which refers to the main steps in any ML pipeline [31]. As for the knowledge representation forms, we relied on the existing taxonomy, although it remains to distill, which forms are actually used in the ML models for KT, and the possibility of expanding the initial list if other forms emerge. On the other hand, we immediately perceived the adaptation of the taxonomy for knowledge sources to our context as more delicate. von Rueden et al. [12] proposed three main sources: scientific knowledge, world knowledge, and expert knowledge. According to their definitions, scientific knowledge mainly refers to science, technology, engineering, and mathematics, and it is validated through formal reasoning or scientific experiments. World knowledge alludes to facts from everyday life, which can be validated implicitly by human reasoning based on intuition; they also subsume linguistics as world knowledge, e.g., syntax and semantics of a language. Expert knowledge is common knowledge within a specific experts' community and is mainly validated through a group of experienced specialists.

Following these definitions, scientific knowledge does not apply to KT, while the other two forms fit with the educational context. However, it is sometimes difficult in our selected papers to distinguish whether a knowledge source is the result of general knowledge or is based on an expert-domain learning theory. Furthermore, the classification in the world and expert knowledge is extensive and does not capture some specificity of the sector, which a finer granularity of the taxonomy might capture. This refinement process was inspired by another taxonomy source borrowed from ITS and AEHS. These systems have four major components [32]: domain model, pedagogical model, LM, and tutor–learner interface model. The latter is mainly a model on a technical level: it determines the admissible inputs (e.g., click,

TABLE II
REFERENCES CLASSIFIED BY KNOWLEDGE REPRESENTATION AND KNOWLEDGE SOURCE

Source		Representation				
		Algebraic equations	Simulation results	Knowledge graphs	Probabilistic relations	Other data
Domain knowledge	Items difficulty	[28], [26], [33], [34], [35], [36], [37], [38], [39]		[40]	[41]	[42]
	Semantic similarity		[27], [43], [44], [45], [46], [47]	[48]		
	Knowledge structure	[49], [50], [51]	[52], [44], [53], [54], [37], [46], [51]	[26], [48], [33], [55], [56], [57], [58], [59], [60], [61], [62], [50], [49]	[13], [63], [64], [65], [66]	
	Class context			[48]	[48], [67]	
Learning knowledge	Pedagogical assumptions		[68]	[69]	[70]	
	Cognitive theories	[71], [20], [38], [39], [47], [19]	[72]		[59], [41], [73]	
Behavioural knowledge	Time	[74], [75], [76]			[34]	[42], [77]
	Scaffolding interactions	[78]			[66]	[77]
	Attempts			[71]		[20], [42], [77], [79]

typing, speech) and produces output in different formats (e.g., text, diagrams, animations, agents); it shapes the architecture through which data are collected; it mediates the interaction between the learner and the contents. On the other hand, the other three components are models for integrating information on different aspects that influence learning into ITS and AEHS. Hence, they are eligible as possible sources of knowledge specific to our topic.

Operationally, a first screening of the paper was made to build a set of labels suitable for classifying each of the three IML dimensions, taking into account the previous considerations. The most appropriate categories were gradually detected as the papers were analyzed. Once we derived the labels, we homogenized and reorganized them to arrive at a stable classification taxonomy. We conducted a second screening phase with this new label set by classifying the 53 papers. During this classification process, we identified the knowledge source enclosed in the model and how it was represented and integrated. Even more labels for each dimension of the taxonomy can be applied to a single paper if there are more types of knowledge sources or if the authors exploit different representation or integration strategies.

In the following session, we introduce the classification taxonomy obtained from the first qualitative analysis of the papers and the classification results in quantitative terms. Furthermore, we chose one of the papers included in our systematic literature review to show how our taxonomy can be applied to describe the prior knowledge injection flow in a real case study.

IV. Results

A. Taxonomy of IML for KT

Here, we present the result of the qualitative analysis that led to the determination of the reference taxonomy for the selected papers' classification. In illustrating our results, we cite only the most recent paper among those included in the systematic review. The full classification, according to the introduced taxonomy, is offered in Tables II and III. The two tables classify the papers by knowledge representation and (path from) knowledge source and by knowledge representation and (path to) knowledge integration.

B. Knowledge Source

The first focus in our taxonomy concerns pointing up the knowledge sources, which can be considered when dealing with KT, i.e., other information retained valuable to integrate those generally used in the standard KT models, that are, the sequence of students' performances. We developed a two-level classification, which expresses different degrees of granularity, summarized in Fig. 2. At the first level, we have three nodes inspired by the ITS components: domain knowledge (domain models), learning knowledge (pedagogical models), and behavioral knowledge (student knowledge).

With the term *domain knowledge*, we indicate both the disciplinary space, i.e., information related in some way to the content object of the learning, and the context where learning occurs. There are four kinds of information included under this umbrella term: items' difficulty, items' semantic similarity, knowledge structure, and class context. The first one refers to information about the difficulty level that characterizes each item used to track the students' knowledge development. It can be assumed either as an intrinsic property of the item, i.e., the level of difficulty is the same for all the students (e.g., [42]), or as a feature to be modeled properly for each student (e.g., [33]). Semantic similarity indicates the benefit of the items' texts as a source of knowledge. The general objective is to exploit semantic similarities between the exercises to highlight valuable

TABLE III
REFERENCES CLASSIFIED BY KNOWLEDGE REPRESENTATION AND KNOWLEDGE INTEGRATION

Integration	Representation				
	Algebraic equations	Simulation results	Knowledge graphs	Probabilistic relations	Other data
Training data	[20], [74], [34], [36], [75], [37], [38], [39], [76]	[27], [43], [44], [52], [45], [68], [54], [37], [47], [46]	[69], [55], [56], [57], [58], [59], [60], [61], [62], [50]	[34], [63], [59]	[20], [42], [77], [79]
Hypothesis set	[71], [33], [38], [78], [47], [49]	[53], [72], [51]	[71], [26], [48], [33], [49], [40]	[48], [13], [70], [41], [64], [65], [67], [66], [73]	
Learning Algorithm	[28], [26], [35], [50], [19]				

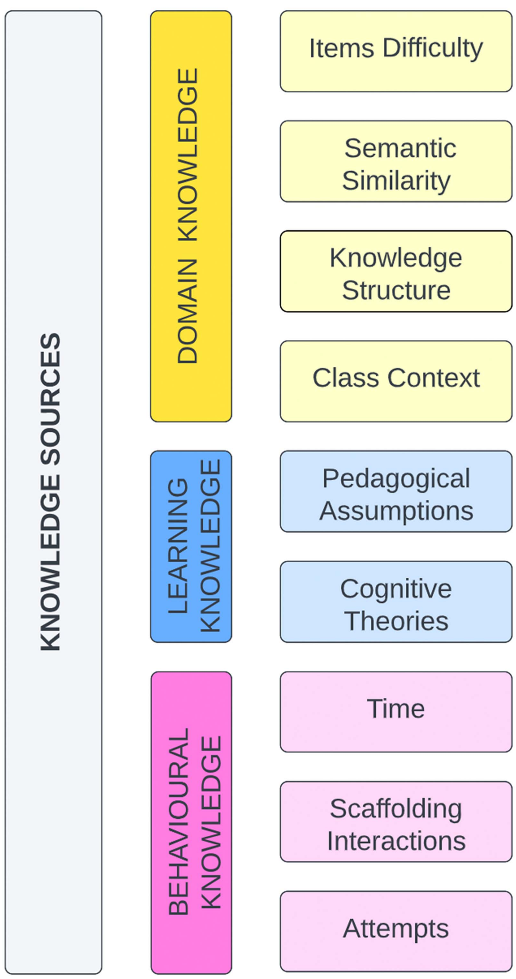

Fig. 2. Taxonomy of knowledge sources for the KT Problem. There are three main classes of knowledge sources for which we have identified some subclasses.

relationships among them, as shown in [48]. In most cases, the integrated knowledge source concerns the knowledge structure, i.e., making explicit the relationships between knowledge concepts, skills, and exercises. This includes both links between concepts or skills considered to be communed by experts (e.g., [56]), and concept(s)-item or skill(s)-item links (e.g., [44]). In this way, we are considering the epistemological structure of a discipline to handle a typical issue in learning assessment: it is not possible to directly measure the students' mastery level of a set of attributes, but it can be inferred by looking for patterns in their items' responses. The last option in this family, namely class context, indicates the use of information about the other students in the class to infer characteristics of the context in which the learning took place, assuming that this can influence each student's learning. For example, Tong et al. [48] considered which exercises are often solved in sequence to infer hierarchical relations between the items. Wang and Beck [67] try to create a model of the class because it can be representative of important information that affects a student's prior knowledge. For example, students in the same class have the same teacher and curriculum and have been assigned the same homework.

The second family of knowledge sources is named *learning knowledge*. It refers to expert knowledge about how learning occurs. We differentiate two types: pedagogical assumptions and cognition theories. In the first group, we enclose theories or hypotheses on learning from an external point of view. For example, Lee et al. [69] cited knowledge space theory as a reference to capture the knowledge structure. They assumed that if students correctly solve a tough exercise on a specific topic, they could even solve correctly other easier exercises on the same topic. Among the cognition theories, we include references to the individual learning process, e.g., the Ebbinghaus forgetting curves proposed in cognitive science studies [71].

As for *behavioral knowledge*, we refer to information concerning how students behave during the learning process in terms of interactions with the learning materials (mainly the items in a learning system). We see a connection with the LM component of the ITS because this information is related to the student. Still, it enriches the exercise-performance sequence traditionally considered in KT. More granularly, we have identified three sublabels: time, scaffolding interactions, and attempts. In [34], they used the average time of answer to estimate items' difficulty. Moreover, information about time is used to estimate the learner's skill mastery, as shown in [42]. As for scaffolding, in education, it refers to breaking up new concepts so they can be learned more easily. Hence, taking into account scaffolding interactions indicates the willingness to integrate the learners' data with information about how or when they use scaffolding materials during their learning process (e.g., [78]). Finally, considering the learners' attempts means monitoring their actions between two consecutive time steps, i.e., determining the knowledge state in a time step also through the attempts and wins/fails ratios that have occurred (e.g., [20]).

C. Knowledge Representation

As categories to define the forms for the representation of knowledge, we referred to the taxonomy introduced by von Rueden et al. [12] (see Table I). Here, we describe only the forms of knowledge representation found in the papers examined in the review.

In most cases, *algebraic equations* are functions to express a mathematical relationship between the variables and constants used to model the problem. Sometimes, the term algebraic constraints are more appropriate because the knowledge is represented through inequalities to determine a feasible set of values (e.g., [71]).

Simulation results are used to describe the numerical outcome of a computer simulation, intended as a way to approximate a real-world scenario and its possible or desirable evolution. There are two recurring forms in the analyzed papers. First, embedding techniques (often pretraining) to obtain more informative representations from the data. Second, the use of attention mechanisms in neural networks. For example, in [44], the domain information on the exercises is integrated through two simulation processes: a pre-training embedding of their texts to gain semantic knowledge and an attention mechanism to model the relations between the items.

Knowledge graph is a common form of knowledge representation. A graph is a pair (V, E), where V is the set of its vertices (or nodes), which usually describe concepts, and E is the set of its edges, i.e., the abstract relations among them. A common knowledge graph within the KT is the Q-matrix, a binary matrix encoding linkages between test items and concepts or other latent or underlying attributes they deal with [80]. It can be provided by an educational expert (e.g., [59]) or estimated directly in the embedding layer by exploiting graph neural networks, which are neural networks designed to handle data represented through graphs (e.g., [55]).

Another knowledge representation type is *probabilistic relation*. According to von Ruedend et al. [12], the correlation of random variables, their conditional independence or the full description of their joint probability distributions can be assumed as a way to represent prior knowledge. This form of knowledge representation is the milestone of BN models [14], very popular as KT techniques.

We add a new class of knowledge representation, named *other data*. There are some cases in which the integrative knowledge source is expressed directly by the collection of additional data to those usually considered in the KT problem (e.g., [20]). As can be seen in Table II, this is quite common when we aim to integrate information on the learner's behavior during learning.

As a final remark to the results concerning knowledge representation, we highlight that there are four classes in the IML taxonomy (see Table I) never used in our qualitative analysis: differential equations, spatial invariances, logic rules, and human feedback.

D. Knowledge Integration

As for knowledge integration, we found three of the four steps of the ML pipeline [31] in the qualitative analysis of the papers.

Integrating prior knowledge sources in *training data* is intended as acting on the information provided as input to the model. There are several ways this can happen. First, we mention data augmentation. Lee et al. [69], for example, defined synthetic data based on pedagogical rationales to deal with the complexity of knowledge acquisition. Another common integration practice is embedding the data with a feature engineering process. This process can be either expert-driven, e.g., in [20], the authors define correct, and error rates as new features to model the students' learning ability, or data-driven, e.g., in [52], a pretraining embedding architecture is designed to model the knowledge structure in the domain. Lastly, some papers expand the training dataset with new kinds of data. In [27], the author leverage knowledge in other domains, which can be transferred to the KT's domain (discipline) object. In other papers, the training dataset includes behavioral data obtained while tracking the learners' interactions with the learning system (e.g., [71]).

The second step of the ML pipeline where prior knowledge can be integrated is the *hypothesis set*. It can be defined as the set of functions to choose to solve the initial problem. Relying on the notation introduced in Section II, the initial problem is estimating $f : \mathcal{X} \to \{0, 1\}$, and the hypothesis set $\mathcal{H}$ is the set of candidate functions among which to choose $g : \mathcal{X} \to [0, 1]$, as a result of the learning algorithm. It may be, for instance, the set of linear functions, the set of neural networks, or the set of logistic functions. Integrating the prior knowledge in the hypothesis set can be intended as bounding the form of the functions included in $\mathcal{H}$. For example, Liu et al. [71] managed two explicit choices in this direction: they exploit recursive functions in their architecture to handle the knowledge master degree estimation according to constructive learning theories; they define a graph convolutional network to include latent learning ability estimation influence on the learner's knowledge concepts states.

The last knowledge integration type found in our literature review is in the *learning algorithm*, i.e., how the model updates the parameters, which define the functions in $\mathcal{H}$ during the training. In a neural networks-based model, this integration consists of modifying the loss function to force the model to consider a prior knowledge source. The authors in [28], for example, introduced a penalization term in the loss function to handle item difficulty.

It is worth noting that we do not have any models where knowledge integration occurs in the final hypothesis step. This kind of knowledge integration would occur when the output of the machine learning pipeline is validated against existing knowledge.

The distilled taxonomy of knowledge representation and integration for KT is summarized in Fig. 3.

E. Quantitative Analysis

We applied the taxonomy described in the previous section to the 53 eligible papers selected for our systematic literature review. One paper can have more than one path for prior knowledge integration. A path is defined by a triad "knowledge source-knowledge representation-knowledge integration." We counted each path separately for the quantitative analysis, identifying

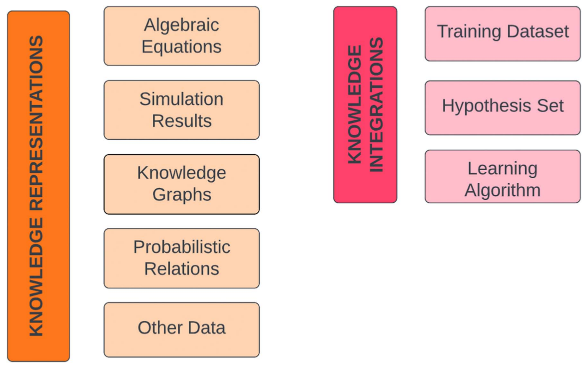

Fig. 3. Taxonomy of knowledge representation and integration for the KT Problem. Five main forms of knowledge representation and three integration steps in the machine learning pipeline are distilled from our systematic literature survey.

77 paths. For instance, Tong et al. [48] introduced 4 ways to integrate exercises learning dependencies in their KT model. As a knowledge source, they exploit knowledge structure from experts, the semantic similarity between items, class context leveraging common behaviors in the class, or class context retrieved by studying the correlation among items answered correctly by many students in the class. They represent the first three knowledge sources using knowledge graphs, while the last is tackled through probabilistic relations. In each path, the integration occurs in the hypothesis set step. Hence, we have 4 triads and considered 4 paths in our labeling counting process.

The quantitative overview of this analysis is summarized through the Sankey diagram in Fig. 4. This visualization format depicts a flow from one set of nodes to another. The paths connect the elements across the three dimensions of the taxonomy and illustrate the approaches we found in the analyzed papers. The height of each node is proportionate to its absolute frequency, which is also expressed in number. The thickness of the links depends on the absolute frequency with which the path connecting two nodes has been recorded.

Let us point out three main pieces of evidence from the Sankey diagram. First, domain knowledge sources are the most exploited, with 40 paths out of 77 which use them. Specifically, in 28 cases, the prior knowledge to integrate concerns the knowledge structure, i.e., more than a third of the paths.

Second, the training dataset is the privileged integration path (43 cases out of 77) for all forms of representation except for probabilistic relations. The latter case is often connected to the choice of the BN as inspiring architecture for the model. Therefore, it is brought back to the hypothesis set class, i.e., the form chosen for the objective function of the predictor.

Third, we have some cases of exclusive inbound or outbound paths. As expected, when knowledge is represented by exploiting other data, we have a unique outbound path to the training dataset, i.e., these new data sources are used to increase the features considered for a more rich knowledge representation. Moreover, the knowledge integration into the learning algorithm step occurs only with an inbound path from algebraic forms of knowledge representation. All the papers that present this approach introduce a regularization term in the loss function, which is optimized during the model training phase.

In addition, the diagram suggests that some representation and integration approaches (paths from knowledge sources to knowledge representations and paths from knowledge representations to knowledge integrations) are more frequent than others, i.e., some paths are more common. A measure for the relevance of each approach is expressed by computing its conditional probability, i.e., the probability that a path ends in a certain node B knowing that the source node is A. In other words, we aim to interpret the Sankey diagram in Fig. 4 as a weighted 3-partite direct graph. The weights of the links are defined through conditional probabilities; the three sets of independent nodes correspond to the three dimensions of the taxonomy (knowledge sources, knowledge representations, and knowledge integrations); the paths' direction is from left to right.

We define the conditional probability $p(B = b_j|A = a_i)$ as

$$p(B = b_j|A = a_i) = \frac{f_{ij}}{f_{i\cdot}} \tag{1}$$

where A and B are two variables, a_i and b_j stand for one of the modalities, respectively, of A and B, f_{ij} is the absolute bivariate frequency (i.e., how many times a_i and b_j occur together), and $f_{i\cdot}$ is the absolute marginal frequency (i.e., the number of total occurrences of a_i). The weights determine the relevance of the different approaches.

For instance, we assume A as the variable for knowledge sources and B as the variable for knowledge integration; $A \in \{a_1, a_2, \ldots, a_9\}$, where a_i denotes the ith modality for knowledge source from the top in Fig. 4; similarly, $B \in \{b_1, b_2, \ldots, b_5\}$, b_j is the jth modality for knowledge representation from the top in the same figure. Hence, we have $a_1 =$ "items difficulty," $b_1 =$ "algebraic equations." The co-occurencies of a_1 and b_1 is $f_{11} = 9$; the marginal frequency for $a_1 n$ is $f_{1\cdot} = 12$. Thus, according to definition 1, we have $p(B = b_1|A = a_1) = 0.75$. In practice, 75% of the paths outgoing from the item difficulty source are integrated into the model through an algebraic representation.

The tables in Fig. 5 summarize the results for all the possible combinations, which define the paths in the Sankey diagram. In particular, we present the contingency tables for the frequencies of each possible path and the adjacency matrices, which describe the graph defined on the Sankey diagram. The elements of the adjacency matrices are the conditional probabilities computed according to definition 1 as weights for the links. More details about the tables are provided in the description of the figure.

From the results presented, we can highlight the following relevant approaches among the paths from knowledge sources to knowledge representations:

1) items difficulty—algebraic equations;
2) semantic similarity—simulation results;
3) cognitive theories—algebraic equations;
4) attempts—other data.

As for the paths from knowledge representations to knowledge integration, the relevant approaches are as follows:

1) probabilistic relations—hypothesis set;

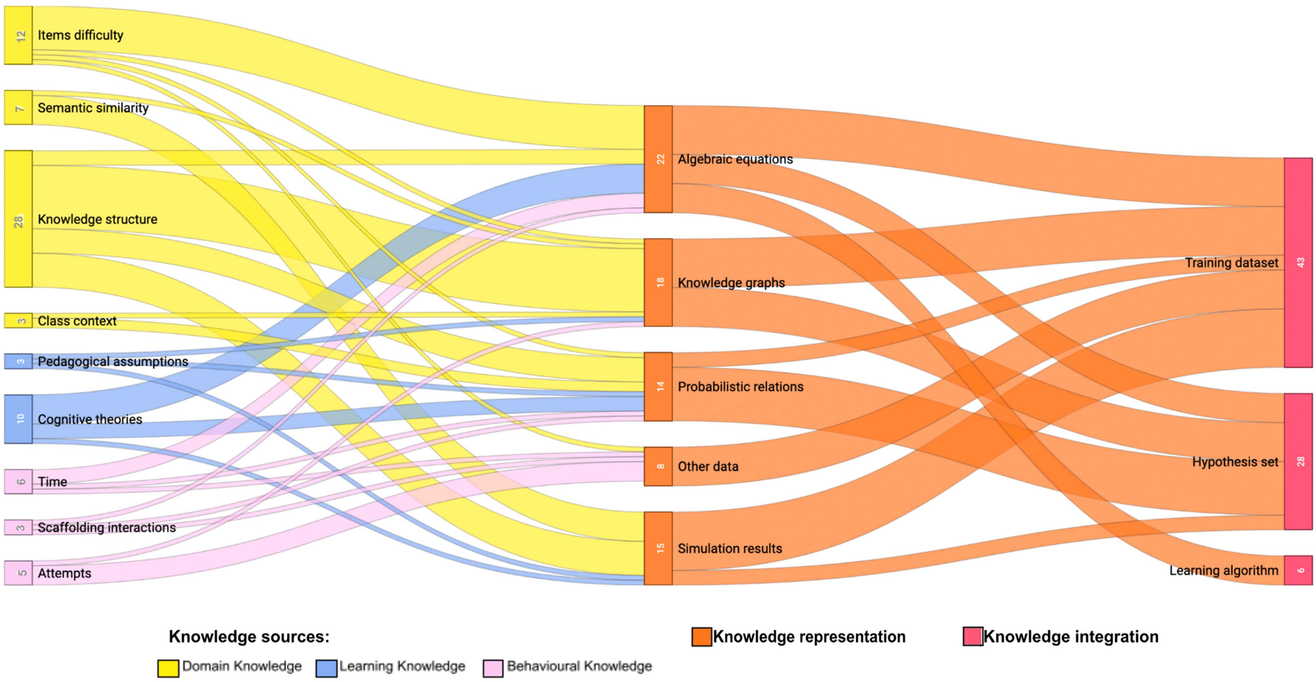

Fig. 4. Paths for integrating prior knowledge in a KT model. The nodes on the left represent the spectrum of knowledge sources distilled from our iterative literature survey; we used three colors to distinguish the three main classes (yellow for domain knowledge, light blue for learning knowledge, and pink for behavioral knowledge). Central nodes cover the forms of knowledge representations. The right nodes are for the types of knowledge integration. The paths among nodes represent different approaches to integrating prior knowledge into the ML pipeline.

	Alg. eq.	Knowl. gr.	Prob. rel.	Other data	Sim. res.	Total
Items diff.	9	1	1	1	0	12
Sem. sim.	0	1	0	0	6	7
Knowl. str.	3	13	5	0	7	28
Class cont.	0	1	2	0	0	3
Ped. ass.	0	1	1	0	1	3
Learn. th.	6	0	3	0	1	10
Time	3	0	1	2	0	6
Scaff. inter.	1	0	1	1	0	3
Att.	0	1	0	4	0	5
Total	22	18	14	8	15	77

(a)

	Alg. eq.	Knowl. gr.	Prob. rel.	Other data	Sim. res.
Items diff.	**0,75**	0,08	0,08	0,08	0,00
Sem. sim.	0,00	0,14	0,00	0,00	**0,86**
Knowl. str.	0,11	0,46	0,18	0,00	0,25
Class cont.	0,00	0,33	0,67	0,00	0,00
Ped. ass.	0,00	0,33	0,33	0,00	0,33
Learn. th.	**0,60**	0,00	0,30	0,00	0,10
Time	0,50	0,00	0,17	0,33	0,00
Scaff. inter.	0,33	0,00	0,33	0,33	0,00
Att.	0,00	0,20	0,00	**0,80**	0,00

(b)

	Train. data.	Hyp. set	Learn. alg.	Total
Alg. eq.	10	6	6	22
Knowl. gr.	10	8	0	18
Prob. rel.	3	11	0	14
Other data	8	0	0	8
Sim. res.	12	3	0	15
Total	43	28	6	77

(c)

	Train. data.	Hyp. set	Learn. alg.
Alg. eq.	0,45	0,27	0,27
Knowl. gr.	0,56	0,44	0,00
Prob. rel.	0,21	**0,79**	0,00
Other data	**1,00**	0,00	0,00
Sim. res.	**0,80**	0,20	0,00

(d)

Fig. 5. Contingency tables and adjacency matrices of the graph for IML taxonomy for KT. (a) Contingency table for all the possible combinations among knowledge sources and knowledge representations modalities (left paths in Fig. 4). (b) Adjacency matrix of the left fold of the graph defined on the Sankey diagram in Fig. 4. (c) and (d) respectively, the contingency table and the adjacency matrix among knowledge representations and knowledge integrations (right fold in Fig. 4). We have bolded in the adjacency matrices the weights associated with the most relevant approaches in the prior knowledge integration pipeline according to the criteria described in the final part of Section IV.

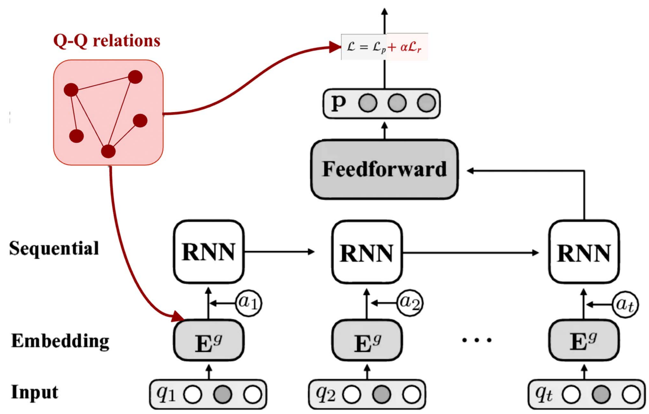

Fig. 6. Architecture for DKT with prior knowledge integration. The figure is adapted from the original paper [50]. In gray, there is the standard architecture for DKT, and in red, the prior knowledge that is integrated. Questions are given as input to an embedding layer. The RNN-based sequential layer is fed with the embedding output and a_i, which encodes whether the student answered the question q_i correctly. Prior knowledge is the similarity relationships between questions. There are two integration paths. First, it is represented through a knowledge graph given as input to the embedding layer. Thus, it is integrated into the training data. Second, it is included in a regularization term added to the binary cross-entry loss. Thus, there is an algebraic reshaping of the learning algorithm.

2) other data—training dataset (already mentioned as an exclusive outbound path);
3) semantic similarity—training dataset.

The relevant approaches have been identified by applying the following criteria: the weight associated with the path is greater than or equal to 60%; both the marginal frequencies and the joint frequency are greater than 5% of the total number of paths, i.e., $f_{ij}, f_{i\cdot}, f_{\cdot j} \geq 4$.

F. Application of IML Taxonomy for KT to a Real Case Study

Before discussing our results for our RQs, we want to show a case of the application of our taxonomy. We aim to make clear what it means to integrate prior knowledge sources in a KT model, following the full path in one of the reviewed papers. We choose the paper by Wang et al. [50] because it is the only one that considers the most frequent category in each dimension of our taxonomy (knowledge structure as knowledge source, algebraic equations as representation form, and training dataset as integration step).

They present a model based on a DKT architecture. In Fig. 6, we present its diagram. The grey part concerns the DKT in its purely data-driven fashion: the student past question–answer sequence feeds an embedding layer; then passes through the sequential layer RNN-based; finally, the feedforward layer predicts the student's future answer to each question.

The red part encodes the prior knowledge injection flows. The first knowledge source consists of question–question relations. These relations are based on their skills and concepts similarity, i.e., two questions are the more similar the more they test the same skills and deal with the same concepts. This knowledge source is represented through a knowledge graph with adjacency matrix A. The adjacency matrix is a square matrix encoding whether pairs of vertices are adjacent or not in the graph. Its integration takes place at the level of training data: the question–question knowledge graph is used as input for the embedding layer of the architecture.

The second knowledge source is the intuition that if a pair of questions requires similar skills or involves similar concepts, students are expected to perform similarly. The knowledge representation here occurs in the form of a regularization term $\mathfrak{L}_{\mathfrak{r}}$, i.e., the algebraic expression $p^T Lp$, where $p(i)$ indicates the probability that the student can answer the ith question correctly and L is the Laplacian matrix associated to A. The loss function in the model is designed to capture this information, and it is defined by two additional terms: $\mathfrak{L}_{\mathfrak{p}}$, the cross-entropy loss, and $\mathfrak{L}_{\mathfrak{r}}$ the regularization term. Thus, the integration occurs at the level of the learning algorithm.

V. Discussion

We now discuss our results, pointing out how they answer to our RQs. We keep two main focuses. First, we highlight some remarks on our distilled taxonomy for IML applied to KT. Given our starting point in von Rueden et al.'s taxonomy [12], it is worthwhile to compare them, to show both points of contact and divergence, and assessing the effectiveness of IML for KT. In this way, we can stress strengths and limitations in our proposal, which can be interpreted as possible future research avenues.

Second, we draw some considerations from the quantitative analysis. We mostly exploited the quantitative results to emphasize relevant and widespread IML approaches among the results. This supports our response to the RQs. We present this discussion in three sections, one for each RQ.

A. Knowledge Sources for KT

As for the knowledge sources, which can be integrated into the ML pipeline to address the KT problem (**RQ1**), we distilled a two-level taxonomy, which is schematized in Fig. 2. Comparing this to the taxonomy for the IML, there is a basic difference in the type of labels that have been searched. In our case, we have identified a label for each type of content or information that is integrated, e.g., under the class "domain knowledge" we have four types of content taken as prior knowledge (i.e., item difficulty, semantic similarity, knowledge structure, and class context). von Rueden et al., on the other hand, defined the labels for the knowledge sources by identifying who holds the prior knowledge, i.e., a scientific theory, a human heritage, or the experts in a specific field.

The strength of our choice is the higher granularity and, thus, its descriptive power. The list of identified classes represents a reference of valuable knowledge sources, which may integrate student performance data in KT tools. They can be considered factors that influence the KT, improving the models' performances; hence, researcher may consider them while developing their models. Also, they can be seen as elements, which enhance the models' explainability and interpretability; that is, they are factors with a high semantic load, making it easier to attribute meaning to the weights or components of the architectures of the designed models (explainability) or favoring the identification of causal relationships between the input and the output of the models (interpretability). In our opinion, this is the first possible line of research that has not yet been sufficiently explored.

On the other hand, the drawback of this improvement in granularity is a loss of generality. Our taxonomy is closely linked to the educational context, with specific reference to the KT problem, while the one for IML applies to domains that may be very different. Furthermore, our taxonomy for knowledge sources is nonexhaustive because it refers to the types of prior knowledge encountered in this literature survey. This does not exclude that there may be other types of relevant content or information to consider. Our result outlines a picture of what exists in the state of the art and can be taken as a starting point subject to updating as a result of new research.

The quantitative analysis highlights one main point: the predominance of domain knowledge as the prior source. Within this class, some subclasses are more exploited than others: knowledge structure is highly considered, while little attention is paid to the class context. The high consideration of the knowledge structure may be due to its easy availability. In fact, in ITS and AEHS, it is often necessary to provide this type of structure to organize the contents of the course. At the same time, the lack of consideration of the class context is not surprising. Most KT systems are applied to online asynchronous learning contexts, where the class context is nonexistent or has little influence. However, it would be interesting, also in light of the teaching experiences that have characterized the recent years of the COVID-19 pandemic, to investigate if and how these systems can be integrated into mixed schooling contexts and estimate which is in this setting the weight of the class context.

As regards other knowledge sources, there are few attempts to consider learning theories, perhaps also due to the difficulty of modeling and representing this kind of knowledge. Most papers in which this occurs refer to cognitive science models of learning curves or forgetting, which are easily representable through algebraic equations. However, there are many other psychological and cognitive factors studied in the literature. For example, a little-considered aspect is the influence of emotions on learning, which has great relevance according to educational experts [81]. This is a further gap to explore in IML for KT.

B. Knowledge Representations for KT

How the prior knowledge is represented (**RQ2**) to attain its integration into the ML pipeline is depicted by the second dimension of our taxonomy (see the left part in the schema of Fig. 3). As already mentioned, we have a subset of the labels used by von Rueden et al. (see Table I), and we add a new label, i.e., other data. The knowledge representations forms never met in our literature survey are differential equations, spatial invariances, logic rules, and human feedback. The first three forms of representation fit better to fields where mathematical modeling of a phenomenon is among the best strategies for its description and study. In the literature survey by von Rueden et al., neither of them is used to represent expert knowledge sources, which is the case for almost all of our knowledge source labels (except semantic similarity, which they include in the class of world or general knowledge). Hence, it is not surprising that they are missing.

On the other hand, we expected human feedback among the knowledge representation forms of our taxonomy. In the IML taxonomy, human feedback "refers to technologies that transform knowledge via direct interfaces between users and machines. [...] Typical modalities include the keyboard, mouse, and touchscreen, followed by speech and computer vision, e.g., tracking devices for motion capturing. [...] This often occurs in areas of reinforcement learning, or interactive learning combined with visual analytics" [12, p.620]. In other words, human feedback, as knowledge representation form, occurs when the human user intuitively and informally expresses a preference or a relevant opinion concerning the output of the automatic model and this is used to enhance the model's performance.

We envision the human feedback representation as useful for personalized learning tools, where KT is the base to suggest to students resources based on their individual needs, thus enabling them to delay or skip contents expected to be too hard or too easy. For instance, camera devices can be exploited to integrate the learner's emotions during the learning process, i.e., the learner facial expressions are assumed in the form of informal human feedback [82]. Another example concerns the interactions among peers in face-to-face lessons, which could be detected by recording audio or asking for explicit feedback from the teacher in the classroom. The main obstacle to this information is the technological equipment normally available to monitor learning, which is connected to a well-known problem of multimodal learning analytics [83]. In other words, this representation also relies on the hardware technology's availability, which affects its effective use. Both examples refer to knowledge sources we have already stated as underconsidered in the papers selected for this systematic literature review. This could justify why human feedback as a knowledge representation form is unused. However, in dealing with a research question about which learning theories may be integrated into KT models, there is an issue on which forms of representations fit better. We believe that an informal representation through human feedback could have an interesting role here to be investigated.

The quantitative analysis has highlighted that the form of representation is often linked to the type of knowledge source:

1) items difficulty and cognitive theories are often represented in algebraic form;
2) knowledge structure is almost represented through graphs;
3) the semantic aspects are represented through simulation results (usually intended as embedding layers, see Section IV, the section on knowledge representation).

Except for the representation form "other data," which is used in only 8 out of 77 paths, the other modalities are distributed evenly. This, in our view, reinforces the need to investigate alternative forms of representation to valorise all knowledge sources that may be relevant to KT.

C. Knowledge Integration for KT

Dealing with RQ3, the labels in our taxonomy are a subset of the one for general IML and are shown in the right part of

Fig. 3. Our literature survey has no cases for integration in the final hypothesis step. The integration in the final hypothesis step occurs when the output of a learning pipeline is "benchmarked or validated against existing knowledge" [12, p.620]. This sounds like ensuring trustworthiness and reliability to the output of the models through a comparison with an authoritative a priori knowledge, i.e., a scientific theory or formal constraints. Such an approach to tackle the KT problem seems unlikely.

The quantitative analysis points out that the training dataset is the privileged step where prior knowledge is integrated into the ML pipeline. In many cases, prior knowledge is used to find effective representations of either the students learning or the items with which they interact. Thus, the tailored representation is a way to arrange differently or augment the training dataset to fully exploit its potential. This trend is not a surprise because it can be led to a characteristic of the phenomenon under study. Specifically, the KT problem is strongly connected with assessing students' learning.

Referring to Højgaard [84, p.4], it is impossible measure students learning directly. Assessment is modeled as a three-step process: "characterizing what you are looking for; identifying the extent to which what you are looking for is present in the situations involved in the assessment; judging the identified." In other words, when dealing with the assessment of students learning, there is an intrinsic problem of identifying some indicators that need to be interpreted in some way. Automating this process, which is usually managed by the teacher, means integrating it into the model. Characterization and identification precede the judgment phase and, in some way, are the premises on which the judgment can be formulated. KT models, according to the definition we presented in Section II, automate the learning judgment phase and use it to predict students' performances on new items. Therefore, studying adequate representations becomes the way to manage the preliminary operation of characterization and identification. It foregoes the model's training, which is oriented to learning how to judge, thus involving mainly the training dataset step.

The quantitative analysis has also brought out three main approaches to knowledge integration based on the knowledge representation kind:

1) the probabilistic relations form is often integrated into the hypothesis set step because probabilistic reasoning is often handled with BNs (chosen as the form for the objective function in the ML pipeline);
2) when the prior knowledge is represented through new data, this is always integrated into the training dataset, becoming an extra input source for the model;
3) simulation results as knowledge representation form is almost integrated into the training dataset.

This last point is in line with two observations already stated in this article: simulation results often occur as embedding layers, thus, connected to the problem of representing the input for the models properly; the representation problem is a necessary pretraining phase, which enable the model to learn how to predict future learners' performances, thus, it is handled in the first phase of the ML pipeline.

VI. Conclusion

To conclude, we summarize the main findings of our systematic literature review and some final remarks.

To answer the three RQs on integrating prior knowledge in KT models, we obtained a three-dimensional taxonomy (knowledge source, knowledge representation, knowledge integration) as main result of a qualitative analysis (see Figs. 2 and 3). This taxonomy has been benchmarked with the one proposed by von Rueden et al. [12] for the IML, taking into account the specific focus on KT. Through a quantitative analysis, some common integration approaches were also identified, which can be deduced interpreting the sankey diagram in Fig. 4. The analysis displays the state of the art at the moment in which the papers involved in the systematic literature review were selected.

Discussing our results in Section V, we have emphasized some gaps in IML for KT, which outline future research directions. We summarize them by posing a new set of research questions (NRQs). They are the result either of strengths that we have found in our taxonomy (e.g., its high granularity), of the assumptions we have made to justify why some prior knowledge injection approaches are more widespread than others, or of some gaps with respect to the literature on learning theories (e.g., the neglect of emotional aspects on KT). In this sense, they represent open issues to be investigated and verified. We formulate the following six questions.

NRQ1 How the integrated prior knowledge sources impact in terms of explainability and interpretability of KT models?

NRQ2 Which prior knowledge sources were not considered in the papers selected for the systematic literature review and could expand the proposed taxonomy?

NRQ3 To what extent KT can be applied in contexts that include face-to-face teaching?

NRQ4 Which role does the class context play as a prior knowledge source in face-to-face (or mixed) teaching settings?

NRQ5 Which cognitive, psychological, or pedagogical theories have relevance in KT (e.g., theory of emotions impact on learning)?

NRQ6 What forms of representation can be used to integrate these theories? (e.g., can we exploit the underconsidered human feedback form for knowledge representation?)

Furthermore, we want to stress that despite the selected papers present hybrid machine learning models, most approaches to KT are still purely data-driven. However, all the papers considered in this systematic literature review claim that their results are comparable to or better than those of traditional ML methods. This encourages further research in this direction. The three RQs posed in Section III can be a trace for researchers to identify, which prior knowledge sources should be considered, how to represent them, and where to integrate them during model development. Our taxonomy can be a tool to use in the exploratory phase to determine what to consider. Moreover, the good performances achieved by these models can be evaluated with respect to the bias issues that characterize AI applications in education [85]. There are different levels at which bias may affect the ML pipeline, e.g., in the data collection process, the data annotation step, the learning algorithm choice, or the performance metrics selection. Integrating prior knowledge can be a new source of bias or, conversely, act as a mitigating effect. This was out of the scope of this work, but future research could investigate the bias challenges of IML for KT.

As a final remark, we point out that in this study we did not consider the implication for practice, i.e., how the results obtained can support teaching and learning. This was outside the aim of the systematic literature review, whose focus was more methodological. However, both to validate the utility of these hybrid machine learning approaches to enhance personalized learning and to investigate some aspects proposed in the NRQs, i.e., models' interpretability or the use of human feedback, it is important to develop research focused on the implication for practice. We explicitly mention the link with the interpretability of the model or the use of human feedback, because these are aspects that directly involve teachers and learners; therefore, a global study of the benefits on teaching and learning is needed.

References

[1] J. R. Anderson, C. F. Boyle, and B. J. Reiser, "Intelligent tutoring systems," *Science*, vol. 228, no. 4698, pp. 456–462, 1985.

[2] V. J. Shute and D. Zapata-Rivera, "Adaptive educational systems," *Adaptive Technol. Training Educ.*, vol. 7, no. 27, pp. 1–35, 2012.

[3] P. A. Cohen, J. A. Kulik, and C.-L. C. Kulik, "Educational outcomes of tutoring: A meta-analysis of findings," *Amer. Educ. Res. J.*, vol. 19, no. 2, pp. 237–248, 1982.

[4] B. S. Bloom, "The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring," *Educ. Researcher*, vol. 13, no. 6, pp. 4–16, 1984.

[5] V. Vagale and L. Niedrite, "Learner model's utilization in the e-learning environments," in *Proc. DB&Local*, 2012, pp. 162–174.

[6] L. D. Kurup, A. Joshi, and N. Shekhokar, "A review on student modeling approaches in its," in *Proc. 3rd Int. Conf. Comput. Sustain. Glob. Develop.*, 2016, pp. 2513–2517.

[7] R. K. Hambleton, H. Swaminathan, and H. J. Rogers, *Fundamentals of Item Response Theory*, vol. 2. New York, NY, USA: Sage, 1991.

[8] J. Templin et al. *Diagnostic Measurement: Theory, Methods, and Applications*. New York, NY, USA: Guilford Press, 2010.

[9] A. T. Corbett and J. R. Anderson, "Knowledge tracing: Modeling the acquisition of procedural knowledge," *User Model. User-Adapted Interact.*, vol. 4, no. 4, pp. 253–278, 1994.

[10] A. Abyaa, M. K. Idrissi, and S. Bennani, "Learner modelling: Systematic review of the literature from the last 5 years," *Educ. Technol. Res. Develop.*, vol. 67, no. 5, pp. 1105–1143, 2019.

[11] S. Minn, "Ai-assisted knowledge assessment techniques for adaptive learning environments," *Comput. Educ.: Artif. Intell.*, vol. 3, 2022, Art. no. 100050.

[12] L. von Rueden et al., "Informed machine learning—A taxonomy and survey of integrating prior knowledge into learning systems," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 1, pp. 614–633, Jan. 2023.

[13] A. Tato and R. Nkambou, "Some improvements of deep knowledge tracing," in *Proc. IEEE 31st Int. Conf. Tools Artif. Intell*2019, pp. 1520–1524. [Online]. Available: https://ieeexplore.ieee.org/document/8995203/

[14] C. Conati, *Bayesian Student Modeling*. Berlin, Germany:Springer, 2010, pp. 281–299, doi: 10.1007/978-3-642-14363-214.

[15] P. Pavlik Jr, H. Cen, and K. Koedinger, "Performance factors analysis—A new alternative to knowledge tracing," *Frontiers*, vol. 200, pp. 531–538, 2009.

[16] H. Cen, K. Koedinger, and B. Junker, "Learning factors analysis—A general method for cognitive model evaluation and improvement," in *Intelligent Tutoring Systems*, M. Ikeda, K. D. Ashley, and T.-W. Chan, Eds. Berlin, Germany: Springer, 2006, pp. 164–175.

[17] C. Piech et al., "Deep knowledge tracing," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, vol. 28, pp. 505–513.

[18] J. Zhang, X. Shi, I. King, and D.-Y. Yeung, "Dynamic key-value memory networks for knowledge tracing," in *Proc. 26th Int. Conf. World Wide Web*, 2017, pp. 765–774.

[19] C.-K. Yeung and D.-Y. Yeung, "Addressing two problems in deep knowledge tracing via prediction-consistent regularization," in *Proc. 5th Annu. ACM Conf. Learn.*, Jun. 2018, pp. 1–10. [Online]. Available: https://dl.acm.org/doi/10.1145/3231644.3231647

[20] X. Sun, X. Zhao, B. Li, Y. Ma, R. Sutcliffe, and J. Feng, "Dynamic key-value memory networks with rich features for knowledge tracing," *IEEE Trans. Cybern.*, vol. 52, no. 8, pp. 8239–8245, Aug. 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9345520/

[21] J. R. Anderson, L. M. Reder, and H. A. Simon, "Situated learning and education," *Educ. Researcher*, vol. 25, no. 4, pp. 5–11, 1996.

[22] R. Nkambou, R. Mizoguchi, and J. Bourdeau, *Advances in Intelligent Tutoring Systems*, vol. 308. Berlin, Germany: Springer, 2010.

[23] J. Vainshtein, R. Esin, and G. Tsibulsky, "Constructing domain model based on logical and epistemological analysis," in *Proc. CEUR Workshop Proc.*, 2020, pp. 140–146.

[24] W. Pietsch, "Aspects of theory-ladenness in data-intensive science," *Philosophy Sci.*, vol. 82, no. 5, pp. 905–916, 2015.

[25] A. Zanellati, S. P. Zingaro, F. Del Bonifro, M. Gabbrielli, O. Levrini, and C. Panciroli, "Informing predictive models against students dropout," *Atti Convegno Nazionale*, pp. 18–25, 2021.

[26] A. Tato and R. Nkambou, "Infusing expert knowledge into a deep neural network using attention mechanism for personalized learning environments," *Front. Artif. Intell.*, vol. 5, Jun. 2022, Art. no. 921476, doi: 10.3389/frai.2022.921476/full.

[27] S. Cheng et al., "AdaptKT: A domain adaptable method for knowledge tracing," in *Proc. 15th ACM Int. Conf. Web Search Data Mining Virtual Event*, Feb. 2022, pp. 123–131. [Online]. Available: https://dl.acm.org/doi/10.1145/3488560.3498379

[28] W. Zhang, K. Qu, Y. Han, and L. Tan, "A novel knowledge tracing model based on collaborative multi-head attention," in *Proc. 6th Int. Conf. Innov. Artif. Intell.*, 2022, pp. 210–215, doi: 10.1145/3529466.3529477.

[29] G. G. Towell and J. W. Shavlik, "Knowledge-based artificial neural networks," *Artif. Intell.*, vol. 70, no. 1/2, pp. 119–165, 1994.

[30] M. J. Page et al., "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," *BMJ*, vol. 372, 2021. [Online]. Available: https://www.bmj.com/content/372/bmj.n71

[31] Y. S. Abu-Mostafa, M. Magdon-Ismail, and H.-T. Lin, *Learning From Data*, vol. 4. New York, NY, USA: AMLBook, 2012.

[32] P. I. Pavlik Jr, K. Brawner, A. Olney, and A. Mitrovic, "Tutoring systems," *Des. Recommendations Intell. Tutoring Syst.: Volume 1-Learner Model.*, vol. 1, pp. 39–68, 2013.

[33] W. Gan, Y. Sun, and Y. Sun, "Knowledge interaction enhanced sequential modeling for interpretable learner knowledge diagnosis in intelligent tutoring systems," *Neurocomputing*, vol. 488, pp. 36–53, Jun. 2022. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0925231222002454

[34] Z. Song, S. Huang, and Y. Zhou, "A deep knowledge tracking model integrating difficulty factors," in *Proc. 2nd Int. Conf. Comput. Data Sci.*, Jan. 2021, pp. 1–5. doi: 10.1145/3448734.3450792.

[35] H. Dai, Y. Zhang, Y. Yun, and X. Shang, "An improved deep model for knowledge tracing and question difficulty discovery," in *PRICAI: Trends in Artificial Intelligence*, vol. 13032, D. N. Pham, T. Theeramunkong, G. Governatori, and F. Liu, Eds. Cham, Switzerland: Springer, 2021, pp. 362–375, doi: 10.1007/978-3-030-89363-728.

[36] L. He, "Integrating performance and side factors into embeddings for deep-learning based knowledge tracing," in *Proc. IEEE Int. Conf. Multimedia Expo.*, Jul. 2021, pp. 1–6. [Online]. Available: https://ieeexplore.ieee.org/document/9428154/

[37] X. Zhang, J. Zhang, N. Lin, and X. Yang, "Sequential self-attentive model for knowledge tracing," in *Artif. Neural Netw. and Mach. Learn.–ICANN*, vol. 12891, I. Farkaš, P. Masulli, S. Otte, and S. Wermter, Eds. Cham, Switzerland: Springer, 2021, pp. 318–330, doi: 10.1007/978-3-030-86362-326.

[38] A. Ghosh, N. Heffernan, and A. S. Lan, "Context-aware attentive knowledge tracing," in *Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2020, pp. 2330–2339. [Online]. Available: https://dl.acm.org/doi/10.1145/3394486.3403282

[39] W. Gan, Y. Sun, X. Peng, and Y. Sun, "Modeling learner's dynamic knowledge construction procedure and cognitive item difficulty for knowledge tracing," *Appl. Intell.*, vol. 50, no. 11, pp. 3894–3912, Nov. 2020, doi: 10.1007/s10489-020-01756-7.

[40] Z. A. Pardos and N. T. Heffernan, "Kt-idem: Introducing item difficulty to the knowledge tracing model," in *Proc. Int. Conf. User Modeling, Adapt., Personalization*, 2011, pp. 243–254.

[41] S. Oeda and K. Asai, "Student modeling method integrating knowledge tracing and IRT with decay effect," in *Proc. EKM, EKAW*, 2016, pp. 19–26.

[42] R. Krishnan, J. Singh, M. Sato, Q. Zhang, and T. Ohkuma, "Incorporating wide context information for deep knowledge tracing using attentional bi-interaction," in *Proc. L2D, WSDM*, 2021, pp. 1–13.

[43] R. Xiao, R. Zheng, Y. Xiao, Y. Zhang, B. Sun, and J. He, "Deep knowledge tracking based on exercise semantic information," in *Emerg. Technol. for Educ.*, vol. 13089, W. Jia et al., Eds. Cham, Switzerland: Springer International Publishing, 2021, pp. 278–289, doi: 10.1007/978-3-030-92836-024.

[44] Q. Liu et al., "EKT: Exercise-aware knowledge tracing for student performance prediction," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 1, pp. 100–115, Jan. 2021. [Online]. Available: https://ieeexplore.ieee.org/document/8744302/

[45] N. Zhang and L. Li, "Knowledge tracing with exercise-enhanced key-value memory networks," in *Knowledge Science, Engineering and Management*, vol. 12815, H. Qiu, H. C. Zhang, Z. Fei, M. Qiu, and S.-Y Kung, Eds. Cham, Switzerland: Springer, 2021, pp. 566–577, doi: 10.1007/978-3-030-82136-446.

[46] H. Tong et al., "Exercise hierarchical feature enhanced knowledge tracing," in *Artificial Intelligence in Education*, vol. 12164, I. I. Bittencourt, M. Cukurova, K. Muldner, R. Luckin, and E. Millán, Eds. Cham, Switzerland: Springer, 2020, pp. 324–328, doi: 10.1007/978-3-030-52240-759.

[47] S. Pandey and J. Srivastava, "RKT: Relation-aware self-attention for knowledge tracing," in *Proc. 29th ACM Int. Conf. Inf. Knowl. Manage. Virtual Event*, 2020, pp. 1205–1214. [Online]. Available: https://dl.acm.org/doi/10.1145/3340531.3411994

[48] H. Tong, Z. Wang, Y. Zhou, S. Tong, W. Han, and Q. Liu, "Introducing problem schema with hierarchical exercise graph for knowledge tracing," in *Proc. 45th Int.ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, Jul. 2022, pp. 405–415. [Online]. Available: https://dl.acm.org/doi/10.1145/3477495.3532004

[49] Y. W. Liang, L. W. Wu, and M. Wang, "Inferring how novice students learn to code: Integrating automated program repair with cognitive model," in *Big Data*, vol. 1120, H. Jin, X. Lin, X. Cheng, X. Shi, N. Xiao, and Y. Huang, Eds. Singapore: Springer, 2019, pp. 46–56, doi: 10.1007/978-981-15-1899-74.

[50] Z. Wang, X. Feng, J. Tang, G. Y. Huang, and Z. Liu, "Deep knowledge tracing with side information," in *Artificial Intelligence in Education*, vol. 11626, S. Isotani, E. Millán, A. Ogan, P. Hastings, B. McLaren, and R. Luckin, Eds. Cham, Switzerland: Springer International Publishing, 2019, pp. 303–308, doi: 10.1007/978-3-030-23207-856.

[51] J. Lee and D.-Y. Yeung, "Knowledge query network for knowledge tracing: How knowledge interacts with skills," in *Proc. 9th Int. Conf. Learn. Anal. Knowl.*, Mar. 2019, pp. 491–500, doi: 10.1145/3303772.3303786.

[52] Y. Ma, P. Han, H. Qiao, C. Cui, Y. Yin, and D. Yu, "SPAKT: A self-supervised Pre-TrAining method for knowledge tracing," *IEEE Access*, vol. 10, pp. 72145–72154, 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9813711/

[53] C. Liu and X. Li, "Memory attentive cognitive diagnosis for student performance prediction," in *Web and Big Data. APWeb-WAIM 2021 International Workshops*, vol. 1505, Y. Gao, A. Liu, X. Tao, and J. Chen, Eds. Singapore: Springer, 2021, pp. 79–90, doi: 10.1007/978-981-16-8143-18.

[54] S. Li, L. Xu, Y. Wang, and L. Xu, "Self-learning tags and hybrid responses for deep knowledge tracing," in *Web Information Systems and Applications*, vol. 12999, C. Xing, X. Fu, Y. Zhang, G. Zhang, and C. Borjigin, Eds. Cham, Switzerland: Springer International Publishing, 2021, pp. 121–132, doi: 10.1007/978-3-030-87571-811.

[55] Y. Yang et al., "GIKT: A graph-based interaction model for knowledge tracing," in *Machine Learning and Knowledge Discovery in Databases*, vol. 12457, F. Hutter, K. Kersting, J. Lijffijt, and I. Valera, Eds. Cham, Switzerland: Springer International Publishing, 2021, pp. 299–315, doi: 10.1007/978-3-030-67658-218.

[56] J. Zhang, Y. Mo, C. Chen, and X. He, "GKT-CD: Make cognitive diagnosis model enhanced by graph-based knowledge tracing," in *Proc. Int. Joint Conf. Neural Netw.*, 2021, pp. 1–8. [Online]. Available: https://ieeexplore.ieee.org/document/9533298/

[57] H. Nakagawa, Y. Iwasawa, and Y. Matsuo, "Graph-based knowledge tracing: Modeling student proficiency using graph neural networks," *Web Intell.*, vol. 19, no. 1/2, pp. 87–102, Dec. 2021.

[58] C. Jiang, W. Ganb, G. Sua, Y. Sunb, and Y. Suna, "Improving knowledge tracing through embedding based on metapath," in *Proc. 29th Int. Conf. Comput. Educ.*, 2021, pp. 11–20.

[59] H. Liu, T. Zhang, F. Li, Y. Gu, and G. Yu, "Tracking knowledge structures and proficiencies of students with learning transfer," *IEEE Access*, vol. 9, pp. 55413–55421, 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9230428/

[60] Y. Sun, L. Wang, Q. Xie, and Y. Dong, "Online programming education modeling and knowledge tracing," in *Knowledge Science, Engineering and Management*, vol. 12274, G. Li, H. T. Shen, Y. Yuan, X. Wang, H. Liu, and X. Zhao, Eds. Cham, Switzerland: Springer International Publishing, 2020, pp. 259–270, doi: 10.1007/978-3-030-55130-823.

[61] S. Tong et al., "Structure-based knowledge tracing: An influence propagation view," in *Proc. IEEE Int. Conf. Data Mining*, 2020, pp. 541–550. [Online]. Available: https://ieeexplore.ieee.org/document/9338285/

[62] F. Ai et al., "Concept-aware deep knowledge tracing and exercise recommendation in an online learning system," *Int. Educ. Data Mining Soc.*, pp. 240–245, 2019.

[63] H. N. M. Ferreira, T. Brant-Ribeiro, R. D. Araujo, F. A. Dorca, and R. G. Cattelan, "An automatic and dynamic knowledge assessment module for adaptive educational systems," in *Proc. IEEE 17th Int. Conf. Adv. Learn. Technol.*, 2017, pp. 517–521. [Online]. Available: https://ieeexplore.ieee.org/document/8001848/

[64] P. Nedungadi and M. S. Remya, "Predicting students' performance on intelligent tutoring system—personalized clustered BKT (PC-BKT) model," in *Proc. IEEE Front. Educ. Conf. Proc.*, 2014, pp. 1–6. [Online]. Available: https://ieeexplore.ieee.org/document/7044200/

[65] T. Käser, S. Klingler, A. G. Schwing, and M. Gross, "Beyond knowledge tracing: Modeling skill topologies with bayesian networks," in *Intelligent Tutoring Systems*, vol. 8474, D. Hutchison et al., Eds. Cham, Switzerland: Springer, 2014, pp. 188–198, doi: 10.1007/978-3-319-07221-023.

[66] M. Sao Pedro, R. Baker, and J. Gobert, "Incorporating scaffolding and tutor context into Bayesian knowledge tracing to predict inquiry skill acquisition," *Educ. Data Mining*, 2013.

[67] Y. Wang and J. Beck, "Class vs student in a Bayesian network student model," in *Artificial Intelligence in Education*, vol. 7926, D. Hutchison et al., Eds. Berlin, Germany: Springer, 2013, pp. 151–160, doi: 10.1007/978-3-642-39112-516.

[68] Q. Pan and T. Tezuka, "Prior knowledge on the dynamics of skill acquisition improves deep knowledge tracing," in *Proc. Proc. 29th Int. Conf. Comput. Educ.*, Bangkok, Thailand, 2021, pp. 22–26.

[69] W. Lee, J. Chun, Y. Lee, K. Park, and S. Park, "Contrastive learning for knowledge tracing," in *Proc. ACM Web Conf.* Virtual Event, Lyon, France, Apr. 2022, pp. 2330–2338, doi: 10.1145/3485447.3512105.

[70] R. S. Baker, S. M. Gowda, and E. Salamin, "Modeling the learning that takes place between online assessments," in *Proc. 26th Int. Conf. Comput. Educ.*, 2018, pp. 21–28.

[71] S. Liu et al., "Ability boosted knowledge tracing," *Inf. Sci.*, vol. 596, pp. 567–587, Jun. 2022. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0020025522001876

[72] T. Long, Y. Liu, J. Shen, W. Zhang, and Y. Yu, "Tracing knowledge state with individual cognition and acquisition estimation," in *Proc. 44th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval. Virtual Event, Canada*, Jul. 2021, pp. 173–182, doi: 10.1145/3404835.3462827.

[73] M. V. Yudelson, K. R. Koedinger, and G. J. Gordon, "Individualized Bayesian knowledge tracing models," in *Artificial Intelligence in Education*, vol. 7926, D. Hutchison et al., Eds. Berlin, Germany: Springer, 2013, pp. 171–180, doi: 10.1007/978-3-642-39112-518.

[74] L. Wei, B. Li, Y. Li, and Y. Zhu, "Time interval aware self -attention approach for knowledge tracing," *Comput. Elect. Eng.*, vol. 102, Sep. 2022, Art. no. 108179. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0045790622004219

[75] D. Shin, Y. Shim, H. Yu, S. Lee, B. Kim, and Y. Choi, "SAINT: Integrating temporal features for EdNet correctness prediction," in *Proc. LAK21: 11th Int. Learn. Anal. Knowl. Conf.*, Irvine, CA, USA, Apr. 2021, pp. 490–496, doi: 10.1145/3448139.3448188.

[76] K. Nagatani, Q. Zhang, M. Sato, Y.-Y. Chen, F. Chen, and T. Ohkuma, "Augmenting knowledge tracing by considering forgetting behavior," in *Proc. World Wide Web Conf.*, San Francisco CA USA, May 2019, pp. 3101–3107, doi: 10.1145/3308558.3313565.

[77] X. Sun, X. Zhao, Y. Ma, X. Yuan, F. He, and J. Feng, "Muti-behavior features based knowledge tracking using decision tree improved DKVMN," in *Proc. ACM Turing Celebration Conf. - China*. Chengdu China, May 2019, pp. 1–6, doi: 10.1145/3321408.3322847.

[78] A. Asselman, M. Khaldi, and S. Aammou, "Evaluating the impact of prior required scaffolding items on the improvement of student performance prediction," *Educ. Inf. Technol.*, vol. 25, no. 4, pp. 3227–3249, Jul. 2020, doi: 10.1007/s10639-019-10077-3.

[79] J.-J. Vie and H. Kashima, "Knowledge tracing machines: Factorization machines for knowledge tracing," in *Proc. AAAI Conf. Artif. Intell.*, 2019, vol. 33, no. 1, pp. 750–757.

[80] M. Birenbaum, A. E. Kelly, and K. K. Tatsuoka, "Diagnosing knowledge states in algebra using the rule-space model," *J. Res. Math. Educ.*, vol. 24, no. 5, pp. 442–459, 1993.
[81] C. M. Tyng, H. U. Amin, M. N. Saad, and A. S. Malik, "The influences of emotion on learning and memory," *Front. Psychol.*, vol. 8, 2017, Art. no. 1454.
[82] F. Böttger et al., "Privacy-preserving and scalable affect detection in online synchronous learning," in *Proc. Eur. Conf. Technol. Enhanced Learn.*, Springer, 2022, pp. 45–58.
[83] P. Blikstein and M. Worsley, "Multimodal learning analytics and education data mining: Using computational technologies to measure complex learning tasks," *J. Learn. Anal.*, vol. 3, no. 2, pp. 220–238, 2016.
[84] T. Højgaard, "Competencies, skills and assessment," in *Proc. Crossing Divides: Proc. 32nd Annu. Conf. Math. Educ. Res. Group Australasia*, Citeseer, 2009, vol. 1, pp. 225–231.
[85] R. S. Baker and A. Hawn, "Algorithmic bias in education," in *Proc. Int. J. Artif. Intell. Educ.*, 2021, pp. 1–41.

Andrea Zanellati received the B.S. and M.S. degrees in mathematics from the University of Bologna, Bologna, Italy, in 2009 and 2011, respectively, where he is currently working toward the Ph.D. degree in data science and computation.

His research interests include the applications of methods and techniques of data science and machine learning in the educational field and large-scale assessment tests.

Daniele Di Mitri received the B.S. degree in computer science from the University of Bari, Bari, Italy, in 2014, the M.S. degree in artificial intelligence from the University of Maastricht, Maastricht, The Netherlands, in 2016, and the Ph.D. degree in learning analytics and wearable-sensors support from Open University, Heerlen, The Netherlands, in 2020.

He is currently a research group leader and PI in Artificial Intelligence in Education with the DIPF Leibniz Institute for Research and Information in Education, Frankfurt, Germany, and a Lecturer with Goethe University, Frankfurt, Germany. His research interests include multimodal learning analytics and artificial intelligence in education.

Maurizio Gabbrielli received the Ph.D. degree in computer science from the Univeristy of Pisa, Pisa, Italy, in 1992.

He was at CWI (Amsterdam), University of Pisa, and University of Udine, Udine, Italy. Since 2021, he has been the Head of Department of Computer Science and Engineering, University of Bologna, Bologna, Italy, where has also been a Full Professor in computer science since 2001. He is also an Associate Dean for AI and digital soul and Director of the Master in Management of Digital Technology, Bologna Business School, and Member of the INRIA project team FOCUS. His research interest include programming languages, constraint programming, and artificial intelligence.

Olivia Levrini received the M.S. and Ph.D. degrees in physics from the University of Bologna, Bologna, Italy, in 1995 and 2000, respectively.

She is currently a Full Professor in physics education and history of physics with the Department of Physics and Astronomy, University of Bologna, Bologna, Italy. Her current research interests include interdisciplinarity in STEM education, educational reconstruction of advanced current topics in physics (thermodynamics, relativityand quantum physics), cognition and conceptual change, identity and processes of appropriation, and instruction design on future-oriented STEM issues (climate change, artificial intelligence, and quantum computing).

Dr. Levrini coordinated the European Erasmus + Projects called I SEE and IDENTITIES, and the Horizon 2020 project called FEDORA. She was the Conference President for the 2019 ESERA conference.

Open Access funding provided by 'Alma Mater Studiorum - Università di Bologna' within the CRUI CARE Agreement