Future Technology
Open Access Journal
ISSN 2832-0379

February 2026| Volume 05 | Issue 01 | Pages 314-323

Journal homepage: https://fupubco.com/futech

https://doi.org/10.55670/fpll.futech.5.1.27

Future Publishing LLC

Article

Artificial intelligence and digital technologies for piano sight-reading skill development: a scoping review

Ruiqing Rui, Muhammad Syawal Amran*, Nurfaradilla Mohamad Nasri

Faculty of Education, University Kebangsaan Malaysia, Selangor, Malaysia

ARTICLE INFO

Article history:
Received 03 September 2025
Received in revised form
05 November 2025
Accepted 07 December 2025

Keywords:
Piano sight-reading, Music pedagogy
Artificial intelligence in education,
Educational technology, Scoping review

*Corresponding author
Email address:
syawal@ukm.edu.my

DOI: 10.55670/fpll.futech.5.1.27

ABSTRACT

Piano sight-reading is a complex cognitive activity that many pupils remain unable to perform despite sustained educational efforts. AI and digital technology have revolutionized numerous educational fields; however, their integration with educational technology for sight-reading piano remains diffuse and concerning to experts due to a lack of coherence across AI-related investigations. This study aims to systematize knowledge on the application of AI and digital technologies in educational technology for sight-reading piano, following the PRISMA-ScR guidelines. A search of four main databases (Web of Science, IEEE Xplore, Scopus, ACM Digital Library) was conducted for papers on AI-related technology for sight-reading piano from 2014 to 2024. This resulted in screening 368 entries to select 33 relevant to the study objective. Five types of technology exist: AI-related intelligent tutoring systems, computer vision and optical music recognition, pattern recognition with deep learning, applications of virtual reality and augmented reality, and mobile and IoT. The study demonstrates a discrepancy between the complexity of AI and accessibility for pupils. AI-powered tutoring systems and deep learning approaches are showing promising results in controlled settings, but evidence on long-term effectiveness remains limited. A fundamental tension exists between analytical sophistication and accessibility: high-performing systems require substantial computational resources, while accessible mobile solutions provide much weaker analytical capabilities. On the other hand, accessibility for pupils remains a top priority, including the use of IoT technology for educational sight-reading piano.

1. Introduction

One of the most complex skills involved in instrumental music learning could be viewed as sight-reading for the piano. This skill encompasses the ability to read and render musical scores accurately upon first viewing. A skilled sight-reader's ability to successfully integrate a number of complex visual-perception skills with bimanual movements and instantaneous musical interpretations can only be described as remarkable [1]. Not only does this skill take a long time to develop for the average student, but a lack of sight-reading ability can continue to pose a challenge for many pianists despite instructors' best efforts to remediate the issue. Studies exploring sight-reading accuracy and a range of variables that can impact that accuracy have found that sight-reading ability encompasses a range of skills that need to be specifically developed [2]. Traditional methods of piano instruction typically address sight-reading only as a secondary issue, incrementally practiced rather than formally instructed. This often takes the form of folk pedagogy, consisting of an increasingly complex repertoire, with the hope that competency can be achieved without specific techniques aimed at developing knowledge of the underlying mental processes that control the activity. A lack of pedagogical materials for sight-reading instruction geared explicitly to that instruction can be noted; the materials that do exist may lack a technology of instruction that directly relates the activity to the mental mechanism [3]. Also, because it is highly labor-intensive, personal instruction at a substantive level can be ruled out for some pupils due to affordability. Current trends in artificial intelligence have driven a fundamental shift across various educational areas. Analysis of AI applications between 2010 and 2020 has documented improvement from simple computer-assisted learning to advanced applications of AI algorithms and computer vision techniques [4]. Technological innovations in AI can align with learning pedagogies by incorporating intelligent learning applications that adapt to learning paths and provide instant learning results and feedback to students [5]. Today, the trend of AI applications continues to witness a fundamental shift due to increased awareness of AI's ability to transform learning applications to meet the specific needs of learning across various disciplines of study [6]. In music education specifically, new technology offers particular opportunities to meet the needs of instrumental instruction. Related applications of virtual and augmented reality have begun investigating immersive methods for learning to play the piano, occupying a space where digital instruction coexists with physical activity [7]. Trends indicate that artificial intelligence can transform learning methodologies by introducing better interface design and more intuitive personalization capabilities [8]. Implementing technology successfully in learning requires understanding good design and developing artificial intelligence literacy to prevent complex technology from hindering educational progress instead of advancing it [9].

Despite the popularity of technology-supported piano learning solutions and applications, existing knowledge on the topic remains fragmented and dispersed over a range of applications and associated technology. Attempts to study a single technology system leave the relevance of other systems, with respect to their relative efficiency and suitability for sight-reading learning, unaddressed. A comprehensive analysis of the range of AI and technology applications for sight-reading piano learning does not exist. This creates a barrier for informed decision-making for technology adopters in the education community and a challenge for researchers to establish promising areas of investigation. This current scoping study aims to bridge this knowledge deficit by methodically surveying AI and technology applications for sight-reading piano instruction. Based on the PRISMA-ScR guideline for conducting a scoping study [10], the current study aims to compile evidence for current technology design efforts and their respective levels of success. This study has three aims: to tabulate existing technology offerings, to survey evidence on current technology levels of pedagogical effect and technical design efforts, and to identify current knowledge gaps for future study.

2. Methods

2.1 Review design

This research used the scoping review methodology to systematically map the digital technologies and artificial intelligence being used for the development of piano sight-reading skills. Scoping reviews are especially suited to areas of novel technology, allowing for the extensive identification and classification of heterogeneous interventions and being amenable to various study designs and outcome measures [11]. A scoping review was preferred over a systematic review because the heterogeneity of technology types, outcome metrics, and study designs in this field precludes meta-analytic synthesis. The method supports exploring the extent of evidence across technology types, application settings, and assessment methods. The process was informed by the PRISMA Extension for Scoping Reviews (PRISMA-ScR) statement for transparency and reproducibility. Although traditional scoping reviews do not exclude studies based on quality, a critical appraisal phase was incorporated because this review aims to inform practice decisions, requiring focus on studies with verifiable technical details [12]. The review aimed to (1) uncover and categorize current AI and digital technologies employed in piano sight-reading education, (2) synthesize evidence on their effectiveness and technical implementation, and (3) determine gaps and future research directions.

2.2 Search strategy

A systematic literature search was conducted in four online databases: Web of Science Core Collection, IEEE Xplore Digital Library, Scopus, and ACM Digital Library. These databases were selected since they comprehensively cover the literature of computer science, engineering, and education technology. Education-specific databases, such as ERIC, are not included, since this review focuses on technical AI implementations rather than general music pedagogy. The search covered publications from January 2014 to December 2024. The starting year was set to 2014 because it coincides with the emergence of deep learning applications in music technology following improvements in convolutional neural networks. Search terms were combined using the Boolean 'AND' and 'OR' operators in groups representing three concepts: (1) AI technology terms, (2) musical instrument terms, and (3) sight-reading instruction terms. Complete search strings used for each database are provided in Table 1. Results were limited to English-language peer-reviewed journal articles and conference proceedings. The reference lists of the included studies were manually checked for additional relevant publications.

2.3 Selection process

The studies were selected based on predefined inclusion and exclusion criteria. The inclusion criteria encompassed the following: (1) publication dates between 2014 and 2024; (2) being peer-reviewed English-language publications; (3) dealing with AI or digital technology for piano sight-reading or piano learning with components concerning sight-reading; and (4) having sufficient detail on technical or empirical levels. For the present review, sight-reading was operationally defined as performing music either at first sight or with minimal prior exposure. These framed studies are concerned with real-time score reading, immediate performance from notation, or technologies designed to facilitate one or both of these skills specifically. Exclusion criteria excluded a study if: (1) it focused exclusively on general piano pedagogy without involvement of technology; (2) it dealt exclusively with non-piano instruments; (3) it was a non-empirical publication that did not present any information about implementation; or (4) the full text was unavailable.

Two reviewers independently screened all the records. Inter-rater reliability was calculated by using Cohen's kappa, yielding $\kappa = 0.88$ for title/abstract screening and $\kappa = 0.85$ for full-text assessment. This reflects almost perfect agreement. Disagreements were resolved through consensus after discussion. Figure 1 illustrates the selection process and its results.

At quality appraisal, studies were assessed using criteria adapted from the Mixed Methods Appraisal Tool (MMAT): (1) methodological rigor, (2) sample adequacy, (3) technical implementation clarity, and (4) relevance to piano sight-reading. Those studies with significant quality concerns or marginal relevance were excluded to ensure the review presents actionable guidance for practitioners.

Table 1. Database search strategies

Database	Search String	Limits Applied	Results
Web of Science Core Collection	TS=("artificial intelligence" OR "machine learning" OR "deep learning" OR "computer vision") AND TS=("piano" OR "keyboard") AND TS=("sight reading" OR "sight-reading" OR "music reading") AND TS=("education" OR "training" OR "learning")	2014-2024; English; Articles & Proceedings	95
IEEE Xplore	("All Metadata":"artificial intelligence" OR "All Metadata":"machine learning" OR "All Metadata":"deep learning" OR "All Metadata":"computer vision") AND ("All Metadata":"piano" OR "All Metadata":"keyboard") AND ("All Metadata":"sight reading" OR "All Metadata":"sight-reading" OR "All Metadata":"music reading") AND ("All Metadata":"education" OR "All Metadata":"training" OR "All Metadata":"learning")	2014-2024; English; Journals & Conferences	138
Scopus	TITLE-ABS-KEY("artificial intelligence" OR "machine learning" OR "deep learning" OR "computer vision") AND TITLE-ABS-KEY("piano" OR "keyboard") AND TITLE-ABS-KEY("education" OR "training" OR "learning") AND TITLE-ABS-KEY("sight reading" OR "sight-reading" OR "music reading")	2014-2024; English; Articles & Conference Papers	93
ACM Digital Library	[All: "artificial intelligence" OR All: "machine learning" OR All: "deep learning" OR All: "computer vision"] AND [All: "piano" OR All: "keyboard"] AND [All: "sight reading" OR All: "sight-reading" OR All: "music reading"] AND [All: "education" OR All: "training" OR All: "learning"]	2014-2024; English; Research Articles	42
Total			368

Note: TS = Topic Search; TITLE-ABS-KEY = Title, Abstract, Keywords; All Metadata/All = Full-text and metadata search. Search conducted in December 2024.

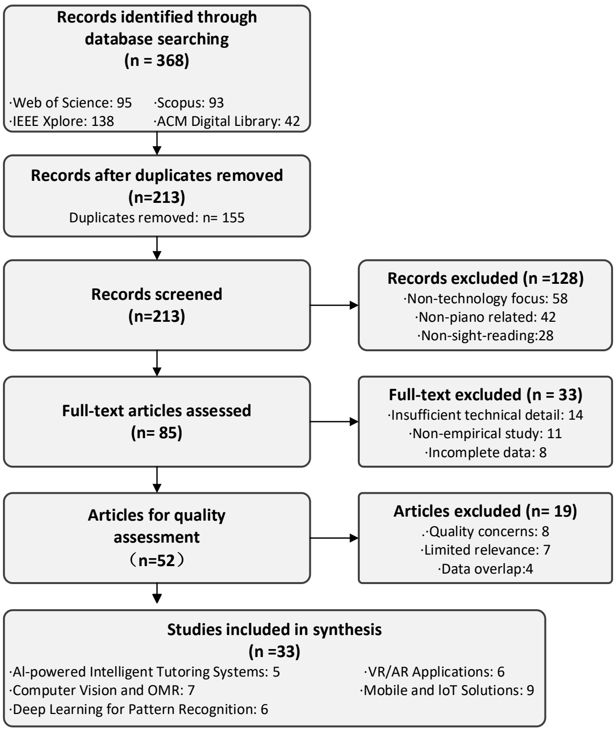

Figure 1. PRISMA flow diagram

2.4 Data synthesis

Data were extracted systematically using a standard template prepared for this review. From each included study, we extracted the following: bibliographic information (authors, year, country), type and category of technology, main algorithms and technical details, study design and methods, sample details, primary findings and results, measures of effectiveness, and limitations noted. Collected data were synthesized using thematic analysis, a continuous cycle of pattern identification, analysis, and reporting across the included studies. Coding followed a hybrid approach: an initial deductive framework based on technology types was applied, followed by inductive refinement as new patterns emerged from the data. Data extraction and coding were managed using Microsoft Excel. Studies were initially coded into five overarching categories of technology types with their main technical focus: (1) AI-based intelligent tutoring systems, (2) computer vision and optical recognition of music, (3) deep learning for pattern recognition, (4) virtual and augmented reality applications, and (5) mobile apps and IoT solutions. Within each category, we used descriptive synthesis to look for common technical characteristics, implementation strategies, and efficacy patterns. Between-category comparisons were then conducted to identify top-level trends, technology convergence, and future development directions. Because of extreme heterogeneity in technology types, study design, and outcome measures, meta-analysis was not feasible; hence, narrative synthesis was used.

3. Results

Studies were categorized by their primary technological approach. The five categories represent distinct technical architectures and pedagogical affordances: AI-powered tutoring (adaptive feedback), OMR (score digitization), deep learning (performance analysis), VR/AR (immersive interaction), and mobile/IoT (accessible delivery). This review comprised 33 studies published from 2015 to 2024, with most (78.8%) since 2020, indicating rapid development in this field. Figure 2 illustrates the temporal distribution of publications, showing a marked increase after 2020 with a peak output in 2022 (n=10). Table 2 presents the distribution characteristics of included studies by publication year, geographic location, study design, and type of technology. Studies came mostly from China (n=12), the United States (n=8), and Europe (n=9), and four from other countries. The evidence pool included empirical research (n=17), technical development articles (n=11), and case studies (n=5). Five technology categories emerged from the analysis: AI-based tutoring systems (n=5), computer vision and optical music recognition (n=7), deep learning for pattern recognition (n=6), VR/AR applications (n=6), and mobile with IoT solutions (n=9). These categories are described in detail in the sections to follow. Table 3 reveals a fundamental trade-off between analytical capability and accessibility across technology types. High-performing systems (AI tutoring, deep learning) face significant computational and cost barriers, while accessible solutions (mobile/IoT) sacrifice analytical depth. This divergence suggests that hybrid approaches combining multiple technology categories may be necessary to balance pedagogical effectiveness with practical implementation constraints.

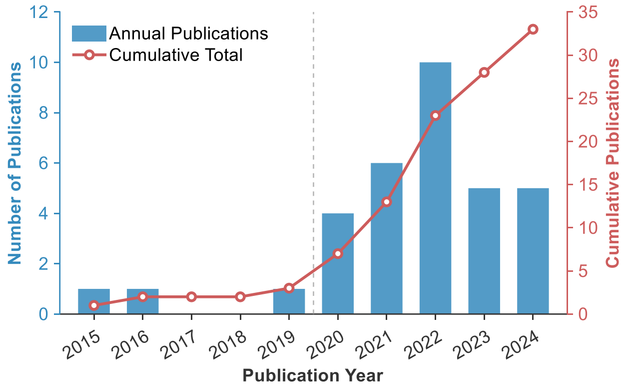

Figure 2. Temporal trends of included studies (2015-2024)

3.1 AI-Powered Intelligent Tutoring Systems

One of the hottest technology domains explored with AI-powered intelligent tutoring systems is artificial intelligence for piano learning tutorials, with machine learning and deep learning techniques being explored to deliver lesson-centric messages with personalized remarks to users. This normally attracts the integration of the implementation of the neural network for determining performance, identifying deficiencies in techniques, and designing roads to learning for the users by assessing their performance [13, 14]. This normally encompasses the performance capture segments, pattern recognition segments, and feedback statements that can be implemented using cloud technology to enable continuous improvement and scaling for better performance [15]. Unlike fixed sets in conventional computer-aided instructional systems, contemporary AI tutoring systems dynamically vary difficulty levels and practice content with respect to individual learning trajectories and performance patterns.

Current implementations exhibit various forms of smart piano pedagogy. Some systems focus on analyzing performance by specifying acoustic and temporal features from student performance and using convolutional neural networks to detect errors in pitch, rhythm, and articulation [13]. Deep learning approaches have been found to be particularly effective at grading fine details of musical expression that are hard for rule-based methods to quantify [14]. Other research features adaptive curriculum sequencing, in which practice history is analyzed using machine learning and used to recommend optimal repertoire-building and technical work [16]. Others involve multimodal analysis of audio recordings, coupled with visual observation of the user's hand positions and posture, to provide comprehensive feedback [17]. By integrating augmented reality into some designs, the tutoring paradigm can expand beyond simple intent analysis to encompass interactive learning that combines digital feedback with actual keyboard learning [17]. All designs are geared toward reducing reliance on one-on-one instruction for comprehensive learning, achieving equal or better learning efficiency, while also decreasing reliance on complete one-on-one instruction to improve learning efficiency. Empirical evidence about learning outcomes remains scant. Li [13] reported a 25-35% improvement in learning efficiency using CNN-based performance analysis, though this was in a small sample in a controlled environment and has not been independently replicated. Other studies have focused mostly on technical accuracy rather than pedagogical outcomes [14,16]. Student engagement seems to improve when gamification elements are included [15]; however, there is a lack of comparative studies across different AI tutoring methods.

Student engagement is also showing a positive trend, particularly when gamification elements are provided to view progress [15]. However, several factors limit the widespread use of the technology. Deep learning models require substantial computational resources and annotated training data, creating barriers for individual learners with limited technical infrastructure [13,14]. Building robust models requires large amounts of annotated piano performances, which remain in short supply in this specialized field [16]. Additionally, concerns about reliance on machine feedback to the exclusion of the development of essential self-assessment skills deserve careful pedagogical consideration [13]. Cost factors also raise issues, as expensive AI systems involve significant development work that may not be accessible within every learning environment.

3.2 Computer vision and optical music recognition

Music recognition technology has also improved immensely over the last few years, from rule-based image processing pipelines to end-to-end deep learning systems. OMR systems aim to convert visual representations of musical notation into machine-readable formats, such as MusicXML or MIDI, enabling digital manipulation, playback, and analysis of handwritten or printed scores [18,19]. In piano sight-reading pedagogy, these technologies perform several tasks: digitization of instructional content for use in interactive learning materials, real-time visual monitoring during practice exercises, and automatic evaluation based on related notes played against familiar score content. Computational difficulty lies in correctly interpreting two-dimensional musical semantics, where pitch and length are encoded by symbol location and morphology rather than a sequential representation [18]. Classic methods used staff detection, symbol breaking, and error-prone classification, while modern methods use integrated neural network models that implicitly learn the music notation hierarchy from examples. Recent technological advances have focused on applying deep learning object-detection methods to music score examination.

Convolutional neural networks have proven robust for note position and duration recognition from score images, and several implementations have tested different network architectures and training schemes [20-22]. Most systems process score pages using hierarchical visual feature extraction across one or more convolutional layers, with classification heads predicting note features such as pitch class, duration, and accidentals. Some methods extend region-based CNN architectures originally designed for overall object detection, treating musical symbols as detection objects in the image of the score [20].

Table 2. Distribution characteristics of included studies (n=33)

Characteristic	Category	Number (n)	Percentage (%)
Publication Year	2015-2017	2	6.1
	2018-2020	5	15.2
	2021-2024	26	78.8
Geographic Origin	China	12	36.4
	United States	8	24.2
	Europe	9	27.3
	Other regions	4	12.1
Study Design	Empirical studies	17	51.5
	Technical development	11	33.3
	Case studies	5	15.2
Technology Type	AI-powered tutoring systems	5	15.2
	Computer vision and OMR	7	21.2
	Deep learning for pattern recognition	6	18.2
	VR/AR applications	6	18.2
	Mobile applications and IoT	9	27.3

Note: OMR = Optical Music Recognition; VR = Virtual Reality; AR = Augmented Reality; IoT = Internet of Things.

Table 3. Cross-category comparison of technology types for piano sight-reading education

Technology Type	Number of Studies	Primary Algorithms	Computational Requirements	Accessibility	Reported Effectiveness	Main Limitations
AI-Powered Intelligent Tutoring Systems	5	CNN, LSTM, Neural Networks, Deep Learning	High	Low	25-35% learning efficiency improvement	High computational costs; requires powerful hardware; large annotated datasets needed; expensive development
Computer Vision and OMR	7	CNN, Transformer, Region-based CNN, Object Detection	Medium-High	Medium	85-95% recognition accuracy	Difficulty with polyphonic scores; challenges with handwritten notation; real-time processing demands
Deep Learning for Pattern Recognition	6	RNN, LSTM, Bi-LSTM, CNN with Attention	High	Low	85%+ agreement with expert evaluation	Large training dataset requirements; model interpretability concerns; limited generalization to unseen repertoire
VR/AR Applications	6	Computer Vision, Hand Tracking, Spatial Computing	High	Low	Enhanced engagement and motivation (not quantified)	High hardware costs; potential for simulator sickness; reduced attention to acoustic output quality
Mobile and IoT Solutions	9	Cloud Computing, Audio Analysis, MSC, QLA	Low (cloud-based)	High	Variable (up to 99%+ accuracy with advanced algorithms)	Weaker analytical capabilities compared to dedicated systems; connectivity dependency; privacy concerns

Note: CNN = Convolutional Neural Network; LSTM = Long Short-Term Memory; RNN = Recurrent Neural Network; OMR = Optical Music Recognition; VR = Virtual Reality; AR = Augmented Reality; IoT = Internet of Things; MSC = Multiple Signal Classification; QLA = Quality-Learning Algorithm.

Others use fully convolutional architectures, which produce dense predictions across the entire image in a single pass [22]. More recently, transformer models have been at the forefront for their ability to represent long-range dependencies in musical sequences and to tackle OMR as a sequence-to-sequence translation problem from image patches to symbolic music notation [23,24]. These are found to hold particular promise for processing polyphonic piano scores with interplaying multiple voices on different staves [23]. Attention Mechanisms help pay heed to significant properties of musical notation while remaining attentive to musical scale, without incurring the costs of full convolutions.

Performance varies significantly depending on the evaluation metrics and datasets. The CNN-based approaches achieve between 85-95% symbol-level accuracy for standard benchmarks like MUSCIMA++ and PrIMuS, which are mainly formed by monophonic or simple polyphonic scores [18, 20]. Transformer-based models exhibit advantages in dealing with longer musical sequences and complex polyphonic textures but usually require substantially bigger training datasets and computational resources [23, 24]. Nevertheless, the tasks of evaluation on piano-specific polyphonic scores have remained scant, and the accuracy usually drops when dealing with complex multi-voice piano repertoires [21].

Polyphonic piano music with complex notation, multiple dynamic markings, and performance marks is considerably more challenging than simple monophonic melodies [21]. Handwritten scores and degraded historic documents introduce additional challenges for identification, necessitating high-quality training data and model adaptation [19]. Real-time requirements for processing interactive training programs offload the burden on efficient inference, necessitating compromises in model complexity and computational tractability [20]. Furthermore, the process of converting recognized symbols into semantically rich musical forms with particular attention to voice leading and harmonic structure remains a research issue to date [24].

3.3 Deep learning for pattern recognition

Deep learning methods are particularly powerful analytical techniques for piano performance analysis, enabling beyond-error analysis of simple patterns and detailed commentary on musicality and technique. Neural network learning with hierarchical representation learning from performance data enables automated analysis of characteristics that, without expert knowledge, only human analysis could correctly determine [25,26]. Analysis power encompasses all levels of analysis performance, from correctness analysis for rhythm precision to pitch-precision analysis, including control of dynamics and temporal synchronization of the hands and fingers. First of all, the key benefit of using deep learning analysis is that patterns can be extracted at a high level of performance complexity without specifically engineering performance characteristics. This works magnificently for catching musical interpretations at a detailed level of performance that rule-description analysis can hardly catch. Different analysis techniques have been implemented using neural network architectures tailored to the specific level of analysis to be carried out.

Recurrent neural networks and their variants, such as Long Short-Term Memory networks, have worked well in describing temporal dependencies in musical performance. Such models depict sequences of performances with recurrently connected representations and internal state representations so that the network can locate each note within its previous musical context. Bi-directional LSTM architectures realize this potential through the addition of new as well as old context, particularly beneficial in error detection, where lack of fit to expected patterns is made apparent through temporal discontinuity [27]. Convolutional neural networks also possess complementary capacity for analyzing spatial patterns within spectrograms or piano roll representations, with recent studies investigating attention mechanisms that enable models to concentrate on musically significant areas when assessing [28]. Dynamic time warping algorithms combined with deep learning frameworks enable comparison between performed and reference renderings even with natural tempo changes [25]. Certain research has tried to push pattern recognition to multimodal analysis, combining audio features with physiological measures like EEG for performance error detection and cognitive load [27]. Such multimodal research points toward the potential for adaptive systems not just to be acoustic output-sensitive but even to performers' mental states. Performance assessment research indicates that the accuracy of deep learning models can be comparable to human expert agreement. Wang and Mukaidani reported an agreement of 85% using DTW-based evaluation, but this was tested on a limited repertoire of classical pieces. Current state-of-the-art models, including Onsets and Frames and transformer-based architectures, have advanced automatic piano transcription, but their application to pedagogical assessment remains underexplored. The most important limitation across studies is the lack of standardized datasets: most models are trained on small, proprietary collections with varying annotation methods, which limits cross-study comparability and generalization to diverse repertoire [29,30].

However, there are certain limitations in existing deployments. Well-trained evaluative models require very large datasets with diverse skill levels, musical styles, and repertoire. The currently available datasets differ significantly in size (ranging from hundreds to tens of thousands of performances), annotation granularity, and genre representation, making direct comparisons of model performance difficult [29]. Moreover, models trained on a particular repertoire often cannot generalize well to unknown musical pieces [30]. Also, the lack of transparency in Deep Neural Networks raises concerns about interpretability, since users may not understand the rationale for a particular performance being generated by them [28]. Lastly, the issue of dataset influence can lead to giving some performance characteristics precedence over others that may be equally fair and valid.

3.4 Virtual and augmented reality applications

Virtual and augmented reality technologies offer distinctive pedagogical affordances for piano instruction by enabling immersive learning environments that combine digital guidance with physical practice. These systems tend to be based on the use of head-mounted displays or spatial computing devices to overlay instructional content on the learner's field of view, providing immediate visual feedback on finger placement, posture, and score interpretation [31,32]. Unlike conventional screen-based lessons that require divided attention between the keyboard and screen, AR applications retain visual attention on the instrument itself by projecting notation, finger numbers, or colored guides onto piano keys [33,34]. VR implementations do it differently, constructing entirely virtual practice spaces where pupils practice with virtual pianos using hand tracking or haptic controllers. Mixed reality configurations combine elements of both paradigms so that real pianos can be observed, with overlays of virtual instructional data or avatars of distant teachers superimposed upon them [31]. The spatial nature of these technologies enables three-dimensional visualization of musical conceptions that are difficult to convey in traditional two-dimensional media, such as hand motion paths and geometric relationships within chord structures. Existing implementations demonstrate varied pedagogical strategies utilizing immersive technologies.

Some of them incorporate gamification techniques in which musical notes stream on the keyboard rhythmically, in a rhythm game fashion, making learning and practicing quite delightful for young learners [35]. Others are focused on developing techniques with continuous visual feedback on hand position and finger form to counterbalance posture-related problems that develop during remote learning by individual students [34]. Some applications incorporate a social learning interface that facilitates remote learning, with the teacher serving as a virtual participant in the learner's mixed reality perception [31]. More sophisticated applications incorporate multimodal analysis that goes beyond computer vision techniques by combining computer vision technology with EMG sensors to analyze muscle activity patterns, with a view to understanding physical tension that can lead to physical injuries [36]. The included six VR/AR studies fall into three categories: usability studies of hardware interfaces [32, 34], technical development of tracking systems [36], and initial pedagogical explorations [31,33,35]. Although these reports increased user engagement and enjoyment, such hedonic outcomes are to be distinguished from pedagogical effectiveness. Notably, none of the included studies used any sight-reading assessment instruments that have seen validation in the literature (e.g., Watkins-Farnum), and hence, there is a difficulty in saying whether immersive technologies enhance actual sight-reading skills or simply add to subjective experience. This technology seems more suited to novice learners, who need gamified, visually engaging practice environments.

However, some limitations exist that may impede its widespread acceptance. The current state of VR/AR technology remains somewhat costly and requires setting up, which may deter non-tech-savvy users. Engaging with the headset for a long time may cause simulator sickness in some users, thereby preventing them from spending a long time learning [32]. This immersion, caused by being secluded from the actual environment, may prevent one from being attentive to acoustic output, while its quality matters for musical performance [31]. Also, the short lifespan of hardware may pose sustainability challenges, as a program intended for current hardware may require an overhaul to run on new hardware.

3.5 Mobile Applications and IoT Solutions

Mobile technology and IoT are perhaps the most democratizing forces in technology-assisted keyboard learning, but they also signal a complete overhaul of the cost and availability of musical instruction. In contrast to specific hardware requirements for AI-assisted instruction or a virtual reality environment for installation, mobile technology leverages the pervasive presence of smartphones and tablets to respond to learning needs with minimal access barriers at all levels of instruction [37,38]. IoT Smart pianos push this paradigm further by incorporating sensors and connectivity into existing pianos, depurposing standard acoustic pianos as data input/output units that can record detailed performance data without necessarily employing audio recordings for learning, using recorders [39,40]. This enables a paradigm shift for learning that happens asynchronously and away from fixed geographic and chronological localities that have hitherto circumscribed musical learning experiences. Its relevance goes beyond simple convenience; a pressing issue of granularity for a technology system, as implemented here, is that it faithfully delivers quality learning material to geographically dispersed populations or learning communities that are simply too poor to afford private educational learning at exorbitantly expensive rates [41]. At the technology system implementation level, there appears to be a mix of learning technologies applied to mobile and IoT applications. First applications of mobile technology tended to centrally involve lesson plans and simple activities that necessarily acted as digital learning notebooks [37]. Additional improvements integrate cutting-edge audio analysis algorithms, enabling smartphones with built-in microphones to provide performance accuracy analysis [42]. Wireless network-based implementations have further enhanced the accuracy of algorithmic performance analysis [43,44].

Cloud architectures integrate to support system operation, allowing intensive computation to be delegated to distant server machines without sacrificing user-friendliness on simple mobile device levels of operation [39]. This IoT system-level approach instead aims to longitudinally analyze the continuous accumulation of learning information input by sensor-laden pianos, uploading learning performance details to analytics databases for longitudinal learning improvement analysis and the discovery of regularly developing performance deficits [40]. Learning improvement algorithms seek to optimize by analyzing trends in musical activities using analytics databases to build customized, smart learning improvement advice based on personal learning patterns [39]. In hybrid online learning, concerning the network accessibility issue, balancing operations to continue running applications locally with occasional synchronization as a necessary condition for running applications when network access is available. The pedagogical implications of mobile and IoT technologies extend beyond technical feasibility to broach underlying questions about the character of musical learning. Through independent practice with immediate feedback, such systems most likely reduce conventional overdependence on regular instructor intervention [42]. With this independence, however, come dangers of reinforcing improper techniques when automated feedback fails to capture nuances of important errors [37]. IoT systems are inherently data-centric, which raises significant privacy and ethical concerns. For example, cloud-based platforms regularly collect data on keystroke-level performance, practice duration, and error patterns- the latter of which can be sensitive when users are minors. Discussion of compliance with data protection regulations such as GDPR or COPPA, as well as considerations of data ownership, retention policies, and third-party sharing practices, is rare within the existing literature [40]. It may also be that the freemium business model prevalent in mobile applications creates unequal access to advanced features, potentially contradicting the democratizing potential of those technologies themselves [37]. Regarding pedagogical outcomes, findings remain fragmented. While studies have indeed shown that real-time audio feedback improves rhythm accuracy [42, 44], others do not measure learning gains but instead focus on system architecture. Comparative studies investigating whether mobile/IoT approaches achieve outcomes at least equivalent to traditional instruction remain absent. There are few long-term efficacy studies, and questions remain about whether mobile-mediated learning builds musical knowledge equivalent to that of traditional instruction [45].

4. Discussion

The present review highlights an underlying tension in technology-assisted piano sight-reading practice: the technology with the greatest analytical capability is far too often inaccessible to those most in need of it. Artificially intelligent learning systems have demonstrated considerable promise in recent applications, with technical accuracy improvements of up to 35% over conventional practice [44]. Such improvements constitute actual pedagogical value. But real-time inference infrastructure within the computational realm faces hurdles that cannot be dismissed as technical [43]. It extends beyond hardware costs to include the lack of large, annotated performance data for strong model training [16]. These results point to a paradox in the contemporary piano sight-reading landscape: the systems that yield the most compelling pedagogical results are precisely those least accessible to learners who might benefit most. Whereas AI-based tutoring systems demonstrate 25-35% efficiency gains, and deep learning-based models achieve expert-level evaluation accuracy, their deployment remains confined to well-resourced institutional settings. Conversely, mobile and IoT solutions attain wide accessibility but at the expense of analytical sophistication. This pattern points to a market failure in educational technology, wherein technological capability and practical utility diverge rather than converge.

Deep-learning approaches to performance evaluation have equaled the performance of expert judgment in a majority of application areas [25], but concerns about model explainability cannot be overlooked. Learners and instructors may not be able to identify the reasoning behind specific assessments being generated [28] and, therefore, may be suspicious of the value of such machine-generated feedback in teaching. Such explainability is particularly required when models automatically identify biases in the training data [29]. Mobile and IoT apps respond to accessibility in different ways, capitalizing on students' existing devices. Cloud architectures unbundled processing loads appropriately [39], enabling real-time feedback in practice problems [42]. The pedagogical trade-offs made here need to be scrutinized in depth. Though such platforms provide equal access to technology-facilitated instruction [41], analysis capability is weaker than that of purpose-designed AI systems. In distant or financially struggling environments where private tutoring in the conventional manner is beyond their budget, such concessions may well be entirely justified [37].

4.1 Practical implications for educators and institutions

Technology selection should be informed by institutional context and learner needs, rather than sophistication per se. Where appropriate, computational infrastructure is in place, AI-driven tutoring systems and deep learning-assisted assessment offer something near to personal feedback; instructors nonetheless have a duty of care to ensure that these augment, rather than replace, human teaching [17, 25]. Computer vision and OMR technologies can aid in the preparation of bespoke learning materials, yet recognition accuracy falls with increasingly complex polyphonic repertoire, necessitating manual checking [18,20]. Immersive VR/AR applications may provide a heightened sense of engagement for beginners. Yet, instructors should emphasize demonstrated pedagogical benefit over entertainment value, given the limited evidence to date regarding actual skill acquisition [31]. In resource-poor and/or dispersed learning settings, mobile and IoT solutions are most accessible [44]. When instructors recommend particular apps, they should prioritize offline functionality, robust feedback mechanisms, and data privacy protections, especially for younger learners [37, 42].

4.2 Research limitations and future directions

This review relies on several methodological limitations. Firstly, the diversity of outcome metrics precluded a quantitative analysis. A good many more concerns were the lack of a longitudinal study. Studies have been conducted for weeks and months; no study was found that investigated retention after the end of training or generalization to an unroutined repertoire. A recent study highlighted the underinvestigated nature of questions about the resilience of long-term skills [41]. Others questioned the ability to assess the value of mobile-mediated learning by reference to musical understanding as provided by conventional learning modalities [45].

Publication bias may overestimate its efficacy because few null results are published. A lack of uncontrolled settings means its efficacy in a more realistic environment has not been adequately explored. Although immersion-related learning-related works have found considerable benefits for user engagement [31], its ability to aid with practicing discipline as a means of developing motivation remains unclear.

Several priority areas for future research emerge from this review. First, there is a need for randomized controlled trials comparing AI-tutoring with traditional instruction over longer periods (e.g., 6-12 months) to determine whether efficiency gains persist beyond initial training. Second, studies should employ validated sight-reading assessments (e.g., Watkins-Farnum) to enable cross-study comparison. Third, work on hybrid systems that integrate multiple technologies (e.g., OMR combined with AR-based finger guidance) may overcome the current trade-off between analytical power and accessibility. Fourth, longitudinal investigations examining skill retention and transfer to unrehearsed repertoire remain notably absent [41,45]. Finally, as these technologies increasingly target younger learners, ethical frameworks addressing data privacy and the appropriate use of AI feedback within formative musical development urgently require attention.

5. Conclusion

This study's scoping review initially probed the landscape of AI and digital technology applications in piano sight-reading instruction to establish that there are five categories of technology, differentiated by pedagogical needs and the specificities of implementation. Through careful aggregation of 33 specific studies published between 2015 and 2024, a remarkable level of advancement in technology applications for sight-reading instruction over the past few years becomes evident, while acknowledging that specific persisting challenges continue to impact the actual implementation of these applications. This information explicitly supports the claim that while there exist specific technology applications that are not adequate to address all needs of sight-reading instruction, pedagogical and technology applications that emphasize sight-reading requirements, AI applications for instruction emphasize complex analysis through significant computation. Similarly, computer vision applications to specific OMR technology signify a lack of adequate musical comprehension. In contrast, applications of AR and VR technology promote immersion, but they entail specific hardware-related costs. In contrast, mobile applications signify a specific level of accessibility, while IoT applications signify a lack of adequate personalization. However, this review has some limitations in its scope. The lack of a quantitative analysis due to the diversity of outcome metrics across studies means that some questions about long-term skill retention remain unanswered, given the relative dominance of short-term outcome assessments. Publication bias may also influence the existing evidence base to some extent, as the vast majority of the literature studied focused on controlled environments rather than actual classroom applications. Long-term learning outcomes may be addressed by future studies that establish a common analytical framework across different outcome studies while exploring hybrid technology solutions that strategically incorporate multiple technology types. With the continued expansion of AI capabilities in multimodal solutions and long-range language modeling, it remains imperative to keep the spotlight on genuine educational needs. Technology should ultimately be viewed as a tool to enhance musical understanding and sight-reading ability, rather than the goal of musical instruction.

Ethical issue

The authors are aware of and comply with best practices in publication ethics, specifically regarding authorship (avoidance of guest authorship), dual submission, manipulation of figures, competing interests, and compliance with research ethics policies. The authors adhere to publication requirements that the submitted work is original and has not been published elsewhere.

Data availability statement

The manuscript contains all the data. However, more data will be available upon request from the authors.

Conflict of interest

The authors declare no potential conflict of interest.

References

[1] Arthur, P., McPhee, E., & Blom, D. (2020). Determining what expert piano sight-readers have in common. Music Education Research, 22(4), 447-456. DOI: 10.1080/14613808.2020.1767559

[2] Mishra, J. (2014). Factors related to sight-reading accuracy: A meta-analysis. Journal of Research in Music Education, 61(4), 452-465. DOI: 10.1177/0022429413508585

[3] Perisynaki, E. (2023). Developing pedagogical material for piano sight-reading, for the context of Greek Conservatoire music education (Doctoral dissertation, University of York).

[4] Zhai, X., Chu, X., Chai, C. S., Jong, M. S. Y., Istenic, A., Spector, M., ... & Li, Y. (2021). A Review of Artificial Intelligence (AI) in Education from 2010 to 2020. Complexity, 2021(1), 8812542. DOI: 10.1155/2021/8812542

[5] Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education–where are the educators?. International journal of educational technology in higher education, 16(1), 1-27. DOI: 10.1186/s41239-019-0171-0

[6] Wang, S., Wang, F., Zhu, Z., Wang, J., Tran, T., & Du, Z. (2024). Artificial intelligence in education: A systematic literature review. Expert Systems with Applications, 252, 124167. https://doi.org/10.1016/j.eswa.2024.124167

[7] Amm, V., Chandran, K., Engeln, L., & McGinity, M. (2024). Mixed reality strategies for piano education. Frontiers in Virtual Reality, 5, 1397154. DOI: 10.3389/frvir.2024.1397154

[8] Chassignol, M., Khoroshavin, A., Klimova, A., & Bilyatdinova, A. (2018). Artificial Intelligence trends in education: a narrative overview. Procedia computer science, 136, 16-24. https://doi.org/10.1016/j.procs.2018.08.233

[9] Walter, Y. (2024). Embracing the future of Artificial Intelligence in the classroom: the relevance of AI literacy, prompt engineering, and critical thinking in modern education. International Journal of Educational Technology in Higher Education, 21(1), 15. https://doi.org/10.1186/s41239-024-00448-3

[10] Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., ... & Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Annals of internal medicine, 169(7), 467-473. DOI: 10.7326/M18-0850

[11] Arksey, H., & O'malley, L. (2005). Scoping studies: towards a methodological framework. International journal of social research methodology, 8(1), 19-32. DOI: 10.1080/1364557032000119616

[12] Peters, M. D., Godfrey, C., McInerney, P., Munn, Z., Tricco, A. C., & Khalil, H. (2020). Scoping reviews. JBI manual for evidence synthesis, 10, 10-46658. https://doi.org/10.46658/JBIMES-20-12

[13] Li, W. (2022). Analysis of piano performance characteristics by deep learning and artificial intelligence and its application in piano teaching. Frontiers in Psychology, 12, 751406. DOI: 10.3389/fpsyg.2021.751406

[14] Liao, Y. (2022). Educational evaluation of piano performance by the deep learning neural network model. Mobile Information Systems, 2022(1), 6975824. https://doi.org/10.1155/2022/6975824

[15] Huang, N., & Ding, X. (2022). Piano music teaching under the background of artificial intelligence. Wireless Communications and Mobile Computing, 2022(1), 5816453. https://doi.org/10.1155/2022/5816453

[16] Chen, C. C., Hung, P., Eğrioğlu, E., & Hsiao, K. L. (Eds.). (2022). Deep Learning in Adaptive Learning: Educational Behavior and Strategy. Frontiers Media SA.

[17] Cui, K. (2023). Artificial intelligence and creativity: piano teaching with augmented reality applications. Interactive Learning Environments, 31(10), 7017-7028. https://doi.org/10.1080/10494820.2022.2059520

[18] Calvo-Zaragoza, J., Jr, J. H., & Pacha, A. (2020). Understanding optical music recognition. ACM Computing Surveys (CSUR), 53(4), 1-35. DOI: 10.1145/3397499

[19] Shatri, E., & Fazekas, G. (2020). Optical music recognition: State of the art and major challenges. arXiv preprint arXiv:2006.07885. https://doi.org/10.48550/arXiv.2006.07885

[20] Lee, M., Kim, H., Moon, M., & Park, S. M. (2021). Computer-Vision-Based Advanced Optical Music Recognition System. Journal of Computational and Theoretical Nanoscience, 18(5), 1345-1351. DOI:10.1166/jctn.2021.9626

[21] Nugroho, D. R., & Zahra, A. (2024). Musical Note Position and Duration Recognition Model in Optical Music Recognition Using Convolutional Neural Network. Journal of Image and Graphics, 12(1). DOI:10.1504/IJART.2021.115764

[22] Andrea, Paoline, & Zahra, A. (2021). Music note position recognition in optical music recognition using convolutional neural network. International Journal of Arts and Technology, 13(1), 45-60. DOI:10.1504/IJART.2021.10035633

[23] Ríos-Vila, A., Calvo-Zaragoza, J., & Paquet, T. (2024, August). Sheet music transformer: End-to-end optical music recognition beyond monophonic transcription. In International Conference on Document Analysis and

Recognition (pp. 20-37). Cham: Springer Nature Switzerland. https://doi.org/10.48550/arXiv.2402.07596

[24] Ríos-Vila, A., Inesta, J. M., & Calvo-Zaragoza, J. (2022, April). On the use of transformers for end-to-end optical music recognition. In Iberian Conference on Pattern Recognition and Image Analysis (pp. 470-481). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-04881-4_37

[25] Wang, A., & Mukaidani, H. (2021, June). An efficient piano performance evaluation model using DTW based on deep learning. In 2021 IEEE 30th International Symposium on Industrial Electronics (ISIE) (pp. 1-6). IEEE. DOI:10.1109/ISIE45552.2021.9576338

[26] Kim, H., Ramoneda, P., Miron, M., & Serra, X. (2022). An overview of automatic piano performance assessment within the music education context. DOI:10.5220/0011137600003182

[27] Ariza, I., Tardón, L. J., Barbancho, A. M., De-Torres, I., & Barbancho, I. (2022). Bi-LSTM neural network for EEG-based error detection in musicians' performance. Biomedical Signal Processing and Control, 78, 103885. https://doi.org/10.1016/j.bspc.2022.103885

[28] Zhao, J., & Yu, K. (2024, September). Recognition and Error Correction of Piano Playing Music based on Spatial Attention Mechanism with Convolutional Neural Network. In 2024 International Conference on Distributed Systems, Computer Networks and Cybersecurity (ICDSCNC) (pp. 1-5). IEEE. DOI: 10.1109/ICDSCNC62492.2024.10941279

[29] Hernandez-Olivan, C., & Beltran, J. R. (2022). Music composition with deep learning: A review. Advances in speech and music technology: computational aspects and applications, 25-50. https://doi.org/10.48550/arXiv.2108.12290

[30] Yang, J., Zhou, Y., & Lu, Y. (2023). Multimedia Identification and Analysis Algorithm of Piano Performance Music Based on Deep Learning. Journal of electrical systems, 19(4). https://doi.org/10.52783/jes.632

[31] Amm, V., Chandran, K., Engeln, L., & McGinity, M. (2024). Mixed reality strategies for piano education. Frontiers in Virtual Reality, 5, 1397154. DOI: 10.3389/frvir.2024.1397154

[32] Rigby, L., Wünsche, B. C., & Shaw, A. (2020, December). piARno-an augmented reality piano tutor. In Proceedings of the 32nd Australian conference on human-computer interaction (pp. 481-491). https://doi.org/10.1145/3441000.3441039

[33] Simion, A., Iftene, A., & Gîfu, D. (2021). An Augmented Reality Piano Learning Tool. RoCHI, 2021, 134-141. DOI:10.37789/rochi.2021.1.1.21

[34] Stanbury, A. J., Said, I., & Kang, H. J. (2021, December). Holokeys: Interactive piano education using augmented reality and iot. In Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology (pp. 1-3). https://doi.org/10.1145/3489849.3489921

[35] Molero, D., Schez-Sobrino, S., Vallejo, D., Glez-Morcillo, C., & Albusac, J. (2021). A novel approach to learning music and piano based on mixed reality and gamification. Multimedia Tools and Applications, 80(1), 165-186. DOI: 10.1007/s11042-020-09678-9

[36] Graf, M., & Barthet, M. (2023). Combining vision and emg-based hand tracking for extended reality musical instruments. arXiv preprint arXiv:2307.10203. https://doi.org/10.48550/arXiv.2307.10203

[37] Ng, S. C., Lui, A. K., & Kwok, A. C. (2015, July). Easy-to-learn piano: A mobile application for learning basic music theory and piano skill. In International Conference on Technology in Education (pp. 103-112). Berlin, Heidelberg: Springer Berlin Heidelberg. DOI:10.1007/978-3-662-48978-9_10

[38] Xia, Y. (2020). Resource scheduling for piano teaching system of internet of things based on mobile edge computing. Computer Communications, 158, 73-84. https://doi.org/10.1016/j.comcom.2020.04.056

[39] Yan, L. (2019). Design of piano teaching system based on internet of things technology. Journal of Intelligent & Fuzzy Systems, 37(5), 5905-5913. DOI:10.3233/JIFS-179172

[40] Fang, B., & Liang, X. (2023). Design and application of cloud computing recommendation based on genetic algorithm in piano online course video system. DOI:10.21203/rs.3.rs-2709888/v1

[41] Ruan, W. (2024). Increasing student motivation to learn the piano using modern digital technologies: independent piano learning with the soft Mozart app. Current Psychology, 43(44), 33998-34008. DOI: 10.1007/s12144-024-06924-3

[42] Yao, Y. (2023). Online and offline hybrid teaching mode of piano education under the background of big data and Internet of Things. Journal of Computational Methods in Science and Engineering, 23(2), 715-724. https://doi.org/10.3233/JCM-226640

[43] Xue, X., & Jia, Z. (2022). The Piano-Assisted Teaching System Based on an Artificial Intelligent Wireless Network. Wireless Communications and Mobile Computing, 2022(1), 5287172. https://doi.org/10.1155/2022/5287172

[44] Li, J. (2022). Study on integration and application of artificial intelligence and wireless network in piano music teaching. Computational Intelligence and Neuroscience, 2022(1), 8745833. https://doi.org/10.1155/2022/8745833

[45] Li, M. (2016). Smart home education and teaching effect of multimedia network teaching platform in piano music education. International Journal of Smart Home, 10(11), 119-132. DOI:10.14257/ijsh.2016.10.11.11