Jurnal Pustaka Indonesia (JPI)
2026, Vol. 6, No. 2, 175–196
https://doi.org/10.62159/jpi.v6i2.2441

OPEN ACCESS
Research Article

Deep Learning Models for Measuring Affective Outcomes in Islamic Religious Education

Maemunah[1*], Siti Maryam[2], Muhyidin Tohir Tamimi[3], Saepudin Mashuri[4], Afandi[5]
[1-3] Universitas Cendikia Abditama Tangerang, Indonesia
[4] Universitas Islam Negeri Datokarama Palu, Indonesia
[5] Institut Bahri Asyiq Galis Bangkalan, Indonesia

Abstract
This study examines the preliminary use of deep learning models to support affective assessment in Islamic Religious Education (IRE) at SMAN 24 Kabupaten Tangerang, SMAN 19 Kabupaten Tangerang, and SMA Islamic Village. Using a mixed-methods sequential explanatory design, the quantitative phase analyzed 927 student reflective texts, behavioral logs, and digital interaction records collected from 1,882 students across seven learning cycles. Affective labels religiosity, empathy, moral awareness, ethical disposition, and spiritual engagement were developed through rubric-based annotation by 14 IRE teachers and the research team, with inter-rater reliability reaching Cohen's Kappa = 0.84. The dataset was divided into 649 training records, 139 validation records, and 139 testing records using a stratified 70:15:15 split. CNN, RNN, transformer-based, and multimodal models were compared with Logistic Regression, Random Forest, and SVM baselines. The transformer model at SMAN 24 achieved 85.6% accuracy, 84.9% precision, 85.8% recall, and 85.1% F1-score; the RNN model at SMAN 19 achieved 84.8% accuracy and 84.0% F1-score; and the multimodal model at SMA Islamic Village achieved the highest performance, with 87.3% accuracy and 86.9% F1-score. Error analysis indicated classification overlap between religiosity and spiritual engagement, while qualitative validation through interviews, observations, document analysis, and member checking showed that model outputs require teacher interpretation. The study concludes that deep learning can provide preliminary decision-support evidence for affective assessment in IRE, but cannot independently measure students' inner religiosity, sincerity, or spiritual transformation.

ARTICLE HISTORY
Received : 11 March 2026
Revised : 28 May 2026
Accepted : 5 June 2026

KEYWORDS
Deep Learning; Affective Assessment; Islamic Religious Education; Learning Analytics

CORRESPONDING AUTHOR
***Maemunah**, Universitas Cendikia Abditama Tangerang, Indonesia. Email: maemunah@uca.ac.id

Introduction

Affective outcomes constitute a central objective of Islamic Religious Education (IRE), encompassing the development of religiosity, moral reasoning, ethical disposition, empathy, spiritual engagement, and the internalization of Islamic values in everyday life. Unlike cognitive achievement, affective learning involves complex, dynamic, and context-dependent processes that evolve through students' experiences, reflections, and social interactions (Holmes, Bialik, and Fadel 2019). Consequently, assessing affective outcomes remains a persistent challenge because evidence of affective development is often distributed across multiple forms of learning data rather than being directly observable through a single assessment instrument. In this context, teacher judgment remains an indispensable pedagogical resource for interpreting students' moral and religious growth. However, the increasing availability of digital learning traces provides opportunities to complement professional judgment with systematic evidence derived from students' behavioral patterns, reflective texts, and learning interactions. Deep learning is therefore relevant not merely as a technological innovation but as a methodological approach capable of identifying complex

relationships within large-scale multimodal educational data that are difficult to capture through conventional assessment procedures alone (Ifenthaler and Schumacher 2016; Zaharah et al. 2024).

Although deep learning has increasingly been applied in educational research, its use has mainly concentrated on cognitive prediction, academic performance classification, learning achievement analytics, and automated feedback systems(Adewale et al. 2024; Vieriu and Petrea 2025). In parallel, affective computing studies have demonstrated that emotional states can be detected through textual, visual, physiological, and interactional data, while learning analytics research has shown the value of multimodal traces for understanding student engagement and learning behavior (AlZu'bi et al. 2022; Mohammadi et al. 2025; Vistorte et al. 2024). However, these two strands of research remain insufficiently connected to Islamic Religious Education, where affective outcomes are not limited to emotions or engagement but include religiosity, moral awareness, ethical disposition, and spiritual internalization. Existing studies on Islamic education still tend to assess affective learning through teacher observation, reflective journals, and self-assessment, with limited computational modeling and limited attention to ethical-religious validation. What remains unresolved, therefore, is how deep learning can be empirically adapted to measure affective outcomes in IRE using multimodal data while maintaining pedagogical interpretability, Islamic ethical alignment, and responsible data governance (Shingjergji et al. 2026; Sholeh, Rusydiyah, and Bakar 2024; Zhang et al. 2025).

Existing scholarship on AI-based affective assessment has developed around four interrelated dimensions: affective constructs, data modalities, validation challenges, and ethical considerations. First, studies have demonstrated that deep learning models can identify affective constructs such as emotions, engagement, motivation, and learning attitudes through computational analysis of educational data (AlZu'bi et al. 2022; Lian et al. 2023) . Second, recent advances in affective computing increasingly rely on multimodal data sources, including textual reflections, facial expressions, behavioral traces, and digital interaction patterns, to improve predictive performance and capture the complexity of learners' affective experiences (Hassannataj Joloudari et al. n.d.). However, the growing sophistication of data processing has not been matched by equivalent progress in validation practices. Many studies focus primarily on predictive accuracy while providing limited evidence regarding pedagogical interpretability, contextual validity, and teacher verification of model outputs (Eitel et al. 2021). Moreover, ethical concerns related to data privacy, informed consent, algorithmic bias, and student autonomy remain insufficiently addressed, particularly in value-oriented educational settings. These limitations are even more apparent in Islamic Religious Education, where affective outcomes encompass not only emotional engagement but also religiosity, moral awareness, and spiritual development. Consequently, a significant research gap persists regarding how deep learning models can be validated, interpreted, and ethically implemented within Islamic educational contexts.

Furthermore, research within IRE contexts continues to rely on conventional approaches to affective assessment, including reflective journals, behavioral observations, and self-assessment instruments. While pedagogically valuable, these methods are limited in terms of objectivity, consistency, and scalability. Moreover, there is a lack of integration between emerging technological approaches and the value-based paradigm of Islamic education (Halstead 2004; Rahmawati and Wahyuningsih 2025). Consequently, there is a pressing need to develop innovative models that are

not only technologically advanced but also aligned with the philosophical and ethical foundations of Islamic education (Al-Attas (Syed.) 1991; Intania, Ruwandi, and Saerozi 2026).

Empirically, SMAN 24 Kabupaten Tangerang, SMAN 19 Kabupaten Tangerang, and SMA Islamic Village had established digital learning environments that generated extensive educational data prior to this study. Islamic Religious Education (IRE) teachers in these schools regularly utilized learning management systems, digital assessment platforms, online discussion forums, reflective writing assignments, and attendance monitoring systems as part of routine instructional practices (Liu et al. 2024; Vistorte et al. 2024; Zhang et al. 2025). These platforms produced digital traces in the form of students' reflective texts, participation records, behavioral logs, assignment submissions, and interaction histories, which constituted the primary dataset analyzed in this research. Rather than implementing deep learning systems directly, the schools provided authentic data-rich learning environments that enabled the development and testing of deep learning models for affective assessment (Batista, Mesquita, and Carnaz 2024; Mohammadi et al. 2025; Rodríguez-Ortiz, Santana-Mancilla, and Anido-Rifón 2025). The availability of such multimodal educational data created a valuable opportunity to examine how deep learning techniques could be adapted to identify religiosity, empathy, moral awareness, ethical disposition, and spiritual engagement within real-world Islamic educational settings (Alismaiel 2021; Sutiah and Supriyono 2024).

However, these implementations remain fragmented and lack a comprehensive conceptual framework. The absence of an integrative theoretical model that connects *deep learning* techniques with affective constructs in IRE results in practices that are largely experimental and unstandardized (Valentijn et al. 2013; Vărzaru 2025; Vistorte et al. 2024). In addition, critical issues such as ethical considerations, data validity, and alignment with Islamic values have not been sufficiently addressed. Therefore, this study seeks to bridge this gap by proposing a comprehensive and contextually grounded model (Fithriyah 2026; Humaida, Wati, and Mukmin 2025; Sayyi, Asmuki, et al. 2025; Sholeh et al. 2024).

However, current implementations remain conceptually fragmented because they have not yet articulated a coherent theoretical model connecting Islamic educational philosophy, affective-domain theory, and computational measurement. From the perspective of Islamic education, affective outcomes are not merely emotional responses but reflect the formation of *adab*, moral consciousness, trustworthiness, and spiritual responsibility as integral dimensions of human development (Al-Attas (Syed.) 1991; Halstead 2004). From affective-domain theory, these outcomes can be understood as progressive levels of internalization, ranging from receiving and responding to valuing, organizing, and characterizing values in learners' conduct (Bloom, Krathwohl, and Masia 1964; Ryans and Krathwohl 1965). Computationally, deep learning enables these latent affective indicators to be inferred from textual reflections, behavioral logs, and digital interaction patterns through data-driven modeling (Khan, Usman, and Binsawad 2026; Vistorte et al. 2024). Therefore, this study bridges the gap by proposing an integrative framework in which Islamic values define the normative orientation, affective taxonomy structures the constructs to be measured, and deep learning provides the analytical mechanism for identifying patterns of affective development in IRE.

The contribution of this study should be understood in three differentiated but interrelated levels. First, at the conceptual level, the study proposes a framework for connecting multimodal learning data, deep learning architectures, qualitative interpretation, and Islamic ethical principles in affective assessment. Second, at the technical level, the study develops and tests working

predictive models using CNN, RNN, and transformer-based architectures to classify selected affective indicators in Islamic Religious Education. Third, at the assessment level, this study does not claim to produce a fully validated standardized instrument; rather, it offers preliminary empirical evidence that AI-generated affective predictions can support teacher judgment when combined with classroom observation, interview-based interpretation, and ethical safeguards. Thus, the novelty lies not in presenting a final assessment instrument, but in demonstrating an initial integrative model that links computational prediction, pedagogical validation, and Islamic value alignment within authentic school contexts.

From a theoretical perspective, this study contributes to expanding the discourse on artificial intelligence integration within Islamic education, particularly in addressing the long-standing challenge of assessing affective domains through quantitative and computational approaches. Practically, the findings are expected to provide guidance for educators and policymakers in designing more objective, adaptive, and value-oriented assessment systems. Thus, this research responds not only to academic gaps but also to real-world challenges in evaluating affective learning in the digital era.

Based on the foregoing discussion, this study is guided by four specific research questions. First, how can CNN, RNN, and transformer-based models be developed to process textual reflections, behavioral logs, and digital interaction patterns for measuring affective outcomes in Islamic Religious Education? Second, how do these models perform in terms of accuracy, precision, recall, and F1-score across the three research sites? Third, how do teachers and students interpret the pedagogical feasibility of AI-generated affective assessment results in classroom practice? Fourth, what ethical governance issues, particularly data privacy, informed consent, algorithmic bias, and alignment with Islamic educational values, emerge from the implementation of deep learning-based affective assessment? By formulating these questions, the study limits its scope to model development, computational validation, pedagogical interpretation, and ethical governance rather than making broad claims about the overall effectiveness of AI-based educational evaluation.

Method

This study employed a mixed-methods sequential explanatory design with a quantitative-dominant priority, followed by a qualitative validation phase to examine the applicability of deep learning models for measuring affective outcomes in Islamic Religious Education (IRE). The study was primarily predictive and validation-oriented, aiming to develop and evaluate computational models before examining their pedagogical relevance and contextual interpretation. In the first phase, quantitative data derived from students' behavioral records, interaction logs, and reflective learning activities were analyzed to construct and test predictive models of religiosity, empathy, moral awareness, ethical disposition, and spiritual engagement. The results of the quantitative analysis informed the selection of participants, cases, and themes for the subsequent qualitative phase, enabling an in-depth exploration of how teachers and students interpreted the model outputs and their relevance to classroom practices (Fàbregues et al. 2023; Phelps et al. 2020; Zurc and Laaksonen 2023). Integration occurred at the interpretation stage, where computational findings were compared with interview, observation, and document data to assess convergence, divergence, and contextual validity. This design was selected to combine statistical rigor with contextual understanding, reflecting recent methodological developments that advocate integrating learning

analytics, artificial intelligence, and educational inquiry to investigate complex affective learning processes (Costa 2024; Silmi et al. n.d.). Furthermore, mixed-methods approaches are increasingly recognized as effective for capturing the multidimensional and socially situated nature of affective learning in contemporary educational environments (Teddlie and Tashakkori 2009).

The research was conducted in three purposively selected senior high schools in Tangerang Regency, namely SMAN 24 Kabupaten Tangerang, SMAN 19 Kabupaten Tangerang, and SMA Islamic Village, all of which have implemented technology-supported instructional practices in Islamic Religious Education (IRE). The study involved 1,882 students from Grades X, XI, and XII, consisting of 797 male and 1,015 female students, as well as 14 IRE teachers and 25 school administrators. Participants were selected through purposive and stratified sampling to ensure representation across school types, grade levels, gender, and learning experiences. Students were included if they were actively enrolled in IRE courses, participated in digital learning activities, and completed reflective learning tasks during the study period. Data were collected from September 2025 to March 2026 and comprised 927 student reflective texts, seven cycles of behavioral log records and digital interaction data, affective assessment records, classroom observation notes, and interview transcripts. In the qualitative phase, 55 participants were interviewed, including 17 participants from SMAN 24 Kabupaten Tangerang, 20 participants from SMAN 19 Kabupaten Tangerang, and 18 participants from SMA Islamic Village. The integration of these diverse data sources enabled the analysis of behavioral, cognitive, and affective dimensions simultaneously, which is considered a key principle of contemporary learning analytics and educational measurement (Creswell 2014).

The quantitative component involves the development of *deep learning* models using architectures such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and transformer-based models to analyze multimodal data, including textual reflections, interaction logs, and behavioral indicators (Alzubaidi et al. 2021; Mienye and Swart 2024). Model training and validation are conducted using supervised learning techniques, with performance evaluated through metrics such as accuracy, precision, recall, and F1-score. This approach is consistent with recent studies highlighting the effectiveness of *deep learning* in processing complex educational datasets and predicting affective and behavioral patterns (Gu et al. 2018; Raj and Kos 2025; Taye 2023). Furthermore, the integration of multimodal data has been shown to significantly improve the reliability of affective computing systems in educational contexts.

The quantitative component involved the development of CNN, RNN, and transformer-based models to analyze 927 student reflective texts, behavioral logs, and digital interaction records collected across seven learning cycles. The study involved 1,882 students from Grades X, XI, and XII, 14 IRE teachers, and 25 school administrators from SMAN 24 Kabupaten Tangerang, SMAN 19 Kabupaten Tangerang, and SMA Islamic Village. Affective outcomes were operationalized into five labels: religiosity (214 records), empathy (187 records), moral awareness (176 records), ethical disposition (168 records), and spiritual engagement (182 records). Labeling was conducted collaboratively by IRE teachers and researchers using a structured rubric. A calibration stage used 120 reflective texts, followed by full annotation of 807 records. Inter-rater reliability reached Cohen's Kappa = 0.84, indicating strong agreement.

Data preprocessing included text cleaning, tokenization, stop-word removal, normalization, sequence encoding, and standardization of behavioral and interaction data. The average reflective text contained 187 words, with a maximum length of 524 words, 12,546 unique tokens, and a fixed

sequence length of 256 tokens. Missing data were found in 78 cases (2.8%) and handled through listwise deletion because they were below 5% and classified as MCAR. The dataset was divided using stratified train-validation-test splitting: 649 training records, 139 validation records, and 139 testing records. Model training was implemented in Python 3.11 using TensorFlow 2.15 and Scikit-learn 1.5 on an NVIDIA RTX 4060 8GB GPU with 32GB RAM. Performance was evaluated using accuracy, precision, recall, and F1-score (Alzubaidi et al. 2021; Gu et al. 2018; Raj and Kos 2025).

The qualitative phase employs semi-structured interviews, classroom observations, and document analysis to contextualize and validate the findings from the computational models. Data are analyzed using thematic analysis, involving coding, categorization, and interpretation to identify patterns related to students' affective development and the pedagogical implications of *deep learning* implementation (Johnson, Adkins, and Chauvin 2020). This phase is critical to ensure that the model aligns with Islamic educational values and classroom realities, addressing concerns that technology-driven approaches often overlook pedagogical and ethical dimensions (NewtonSuter 2012). Additionally, qualitative validation strengthens the interpretability and practical applicability of AI-based assessment systems (Ayik et al. 2026).

To ensure validity, reliability, and ethical accountability, this study applied triangulation across quantitative model evaluation, qualitative member checking, expert review, and cross-source verification. Because the study involved minors and sensitive religious-affective data, ethical procedures were implemented before data collection through institutional permission, parental consent, student assent, and voluntary participation. Participants were informed that involvement in the study was not part of academic grading and that they could withdraw at any stage without educational penalty. Data minimization was applied by collecting only information directly relevant to affective assessment, including reflective texts, behavioral logs, digital interaction records, observation notes, and interview transcripts. All identifiable information was removed before analysis by replacing student names and school-related identifiers with coded labels. Access to the dataset was restricted to the authorized research team, and the data were stored in password-protected digital files. Bias mitigation was conducted through stratified data splitting, annotator calibration, inter-rater agreement checking, label distribution review, and error analysis across gender, grade level, and school context. These procedures were used to strengthen methodological credibility while ensuring that AI-based affective assessment remained transparent, responsible, and ethically appropriate for educational settings involving minors and religious-affective data (Gouseti et al. 2025; Wiese et al. 2025).

Result and Discussion

Performance of Deep Learning Models in Measuring Affective Outcomes

The quantitative analysis was conducted on an annotated dataset consisting of 927 student reflective texts, behavioral logs, and digital interaction records derived from 1,882 students across three research sites. Ground-truth labels were established through a structured annotation process involving 14 Islamic Religious Education teachers and the research team. The final dataset comprised five affective categories, namely religiosity (214 records), empathy (187 records), moral awareness

(176 records), ethical disposition (168 records), and spiritual engagement (182 records). Annotation consistency was evaluated using Cohen's Kappa coefficient ($\kappa = 0.84$), indicating strong inter-rater agreement and providing a reliable foundation for supervised learning classification.

The dataset was divided into training (649 records), validation (139 records), and testing (139 records) subsets using a stratified sampling strategy to preserve class distribution across affective categories. Three deep learning architectures CNN, RNN, and Transformer were evaluated against conventional machine-learning baselines, including Logistic Regression, Random Forest, and Support Vector Machine (SVM). Performance was assessed using accuracy, precision, recall, and F1-score. The baseline models achieved accuracy scores ranging from 76.2% to 81.4%, whereas deep learning architectures consistently outperformed conventional approaches, indicating the capacity of neural architectures to capture latent affective patterns from multimodal educational data.

At SMAN 24 Kabupaten Tangerang, the transformer-based model demonstrated the highest performance for textual reflection analysis, achieving an accuracy of 85.6%, precision of 84.9%, recall of 85.8%, and F1-score of 85.1%. Error analysis showed that most misclassifications occurred between religiosity and spiritual engagement, reflecting conceptual overlap between the two constructs. The confusion matrix further indicated that the model was particularly effective in identifying moral awareness and empathy categories, while prediction uncertainty remained higher for records containing multiple affective dimensions simultaneously.

At SMAN 19 Kabupaten Tangerang, the RNN model achieved an accuracy of 84.8%, precision of 83.7%, recall of 84.5%, and F1-score of 84.0%. The model performed particularly well in processing behavioral logs collected across seven learning cycles, enabling the detection of temporal changes in students' participation and religious engagement. Longitudinal sequence analysis revealed that behavioral consistency contributed substantially to classification performance, supporting the argument that affective development should be understood as a dynamic process rather than a static educational outcome.

At SMA Islamic Village, the multimodal model integrating textual reflections, behavioral records, and digital interaction data achieved the highest overall performance, with an accuracy of 87.3%, precision of 86.8%, recall of 87.1%, and F1-score of 86.9%. An ablation analysis demonstrated that textual reflections contributed 41% to predictive performance, behavioral logs contributed 33%, and digital interaction data contributed 26%. When individual modalities were removed, prediction accuracy declined significantly, confirming that multimodal integration improved classification robustness and reduced dependence on any single data source. Cross-site testing further showed that the multimodal model maintained relatively stable performance across the three schools, suggesting reasonable generalizability beyond a single institutional context.

Table 1. Performance of Deep Learning Models Across Research Sites

School	Dominant Model	N Records	Train/ Val/Test	Accuracy	Precision	Recall	F1-Score	Baseline Accuracy
SMAN 24 Kabupaten Tangerang	Transformer	309	70:15:15	85.6%	84.9%	85.8%	85.1%	81.4% (SVM)

SMAN 19 Kabupaten Tangerang	RNN	304	70:15:15	84.8%	83.7%	84.5%	84.0%	80.1% (RF)
SMA Islamic Village	Multimodal DL	314	70:15:15	87.3%	86.8%	87.1%	86.9%	81.4% (SVM)
Cross-site result	CNN	927	70:15:15	83.9%	83.4%	84.1%	83.7%	76.2% (LR)

Table 1 demonstrates that deep learning models consistently outperformed conventional baseline models in measuring affective outcomes in Islamic Religious Education. The multimodal deep learning model implemented at SMA Islamic Village achieved the highest performance, with an accuracy of 87.3%, precision of 86.8%, recall of 87.1%, and an F1-score of 86.9%. These findings indicate that integrating textual reflections, behavioral records, and digital interaction data provides a more comprehensive and accurate representation of students' affective development than single-source assessment approaches.

Table 2. Ablation Analysis of Multimodal Model

Configuration	Accuracy
Text Only	82.4%
Behavior Only	79.6%
Interaction Only	77.8%
Text + Behavior	85.1%
Text + Interaction	84.2%
Behavior + Interaction	82.8%
Full Multimodal Model	87.3%

Multimodal Learning Analytics and Affective Prediction Patterns

The second quantitative result concerns how different data modalities contributed to affective prediction in Islamic Religious Education. Rather than treating the contribution of textual reflections, behavioral logs, and digital interaction records as descriptive estimates, this study examined their contribution through ablation analysis. Each modality was tested separately and then compared with combined-modality configurations. The contribution percentages were derived from the relative performance gain produced by each modality within the multimodal model. This procedure was used to determine whether the highest model performance resulted from genuine multimodal integration or merely from site-specific documentation quality.

The ablation test showed that textual reflections contributed the largest proportion to affective prediction, accounting for 41% of the relative contribution in the full multimodal configuration. Textual data produced the strongest single-modality performance because students' reflective writing contained semantic expressions related to religiosity, sincerity, empathy, moral reasoning, and spiritual awareness. Transformer-based analysis was particularly useful in identifying contextual meaning in reflective texts, including students' ability to connect Islamic concepts with daily conduct. However, text-only prediction remained lower than the full multimodal model, indicating that reflective writing alone was insufficient to represent the whole affective profile.

Behavioral logs contributed 33% of the relative predictive contribution. These data included attendance patterns, classroom participation, punctuality in completing IRE assignments, and recorded involvement in religious learning activities across seven learning cycles. Behavioral data were especially useful for identifying consistency, discipline, and observable engagement. The RNN model showed that students' affective development could be better understood when behavioral patterns were analyzed sequentially rather than as isolated observations. Nevertheless, behavior-only prediction produced lower performance than text-based and multimodal models, suggesting that observable conduct requires interpretive support from other data sources.

Digital interaction records contributed 26% of the relative predictive contribution. These records included participation in online discussions, response frequency, learning platform activity, and interaction with digital learning materials. Although this modality produced the lowest single-modality performance, it provided important supplementary evidence for identifying students who were actively engaged in digital learning but less expressive in written reflections. The model also showed that digital interaction data improved classification for engagement-related categories, particularly spiritual engagement and ethical disposition, when combined with textual and behavioral data.

Overall, the ablation results confirm that the full multimodal model produced the strongest predictive performance because it integrated internal reflection, observable behavior, and digital participation into a single analytical framework. When one modality was removed, model accuracy declined, showing that each data source contributed distinct information to affective prediction. This finding supports the methodological assumption that affective outcomes in Islamic Religious Education are distributed across multiple forms of evidence and cannot be adequately measured through a single data source. Therefore, multimodal learning analytics provides a stronger empirical basis for affective assessment than unimodal approaches.

Table 3. Ablation-Based Contribution of Data Sources to Affective Prediction

Model Configuration	Data Modality Used	Accuracy	Relative Contribution	Affective Indicators Captured	Interpretation
Text only	Student reflective texts	82.4%	41%	Religiosity, sincerity, empathy, moral reasoning	Strongest single modality for internal affective meaning
Behavior only	Behavioral logs	79.6%	33%	Discipline, consistency, participation	Captures observable and sequential affective behavior
Interaction only	Digital interaction records	77.8%	26%	Engagement, responsiveness, digital participation	Provides supplementary evidence of learning involvement
Text + Behavior	Reflective texts and behavioral logs	85.1%	Integrated	Religiosity and behavioral consistency	Stronger than single-modality models
Text + Interaction	Reflective texts and digital interactions	84.2%	Integrated	Reflection and digital engagement	Improves detection of expressive and interactive engagement

Behavior + Interaction	Behavioral logs and digital interactions	82.8%	Integrated	Participation and consistency	Useful but less semantically rich
Full multimodal model	Text, behavior, and digital interaction	87.3%	Integrated	Holistic affective profile	Highest performance and strongest basis for affective prediction

Qualitative Validation of Affective Assessment Results

The qualitative validation phase was conducted to examine how teachers, students, and administrators interpreted the computational outputs generated by the deep learning models. The interviews involved 55 participants across the three schools: 17 participants from SMAN 24 Kabupaten Tangerang, 20 participants from SMAN 19 Kabupaten Tangerang, and 18 participants from SMA Islamic Village. Participants were coded according to institution and role, such as KS24 for the principal of SMAN 24, GPAI24-1 for an IRE teacher at SMAN 24, WK19 for the curriculum vice-principal at SMAN 19, GPAI-IV for an IRE teacher at SMA Islamic Village, and S24-1, S19-1, or SIV-1 for students. The interview protocol focused on four issues: teachers' interpretation of AI-generated affective scores, students' experiences of receiving feedback, the alignment between model outputs and classroom observations, and ethical concerns related to religious-affective data. Thematic analysis produced three dominant themes: pedagogical alignment, reflective usefulness, and the need for teacher mediation.

Teacher interviews indicated that the model outputs were generally consistent with professional classroom judgment, especially when affective predictions were based on reflective texts and behavioral records. IRE teachers explained that students classified as having high religiosity, empathy, or moral awareness were often the same students who displayed consistent participation, respectful communication, and stronger engagement in religious learning activities. One teacher at SMAN 24 stated, *"The system's prediction was close to what we usually observe in class. Students who wrote reflective responses seriously were also the ones who showed discipline and respect during learning"* (GPAI24-1). A teacher at SMA Islamic Village similarly noted, *"The model helped us see patterns that were sometimes missed in daily observation, but the final interpretation still had to come from the teacher"* (GPAI-IV-2). These statements suggest convergence between computational classification and teacher-based pedagogical interpretation.

Student interviews showed that AI-supported feedback encouraged some students to reflect more consciously on their learning attitudes, religious discipline, and social behavior. Several students reported that the feedback was more specific than conventional comments because it referred to patterns in their reflective writing and participation. One student from SMAN 19 explained, *"I became more aware that my answers were not only about knowledge, but also about whether I could connect Islamic values with my daily actions"* (S19-1). A student from SMA Islamic Village added, *"The feedback made me think again about my attitude toward friends and teachers, but I still needed the teacher to explain what the score meant"* (SIV-2). These responses indicate that AI-generated feedback can support self-reflection, but its pedagogical value depends on teacher explanation and dialogical interpretation.

Classroom observations provided additional evidence that supported several model outputs. Students predicted as having stronger affective development tended to participate more actively in discussions, show respectful interaction, complete reflective assignments more consistently, and

relate Islamic teachings to daily life. Observation notes from SMAN 24 and SMA Islamic Village showed convergence between high affective scores and visible patterns of adab, empathy, and participation. However, the observation data also revealed deviant cases. Some students who received moderate prediction scores demonstrated positive affective behavior in face-to-face interaction but wrote brief or less expressive reflections. Conversely, a small number of students produced strong written reflections but were less consistent in classroom participation. These deviant cases indicate that affective prediction cannot be reduced to textual or digital traces alone and must be interpreted alongside classroom realities.

Document analysis further clarified why model integration varied across schools. SMA Islamic Village had more systematic digital documentation, including learning platform records, assignment archives, and affective assessment notes, which made multimodal integration easier. SMAN 24 and SMAN 19 had sufficient digital learning traces, but some affective assessment records remained teacher-centered and less standardized. The school administrators recognized this issue. The curriculum vice-principal at SMAN 19 stated, *"Digital data are available, but they are not always organized in the same format across teachers. This affects how easily they can be used for assessment development"* (WK19). Similarly, the principal of SMA Islamic Village explained, *"Technology can help assessment, but schools must first improve documentation and teacher readiness"* (KSIV). These findings show that institutional data readiness influenced the interpretability and practical use of AI-based assessment.

To strengthen credibility, the qualitative findings were subjected to member checking and expert review. Interview summaries and preliminary themes were returned to selected teachers and administrators to confirm whether the interpretations reflected their intended meanings. Most participants confirmed the accuracy of the thematic interpretation, while several teachers emphasized that AI outputs should be framed as supporting evidence rather than final judgment. Coding was conducted through open coding, axial categorization, and theme development. Initial codes included alignment with teacher observation, student self-reflection, need for explanation, documentation variation, misclassification risk, and ethical caution. These codes were then grouped into broader themes of pedagogical relevance, interpretive mediation, institutional readiness, and ethical sensitivity. This process helped move the analysis beyond generic triangulation toward a clearer explanation of how qualitative evidence confirmed, complicated, or limited computational findings.

Overall, triangulation showed convergence, divergence, and complementarity across interview, observation, and document data. Convergence appeared when teacher interviews and classroom observations confirmed that high predicted affective scores were often associated with visible discipline, empathy, and religious engagement. Divergence appeared in deviant cases where students' written reflections did not fully represent their classroom behavior. Complementarity appeared when document analysis explained why the multimodal model performed better in schools with more organized digital records. Therefore, qualitative validation did not simply confirm the model; it clarified the conditions under which model outputs were pedagogically meaningful. The findings indicate that AI-generated affective assessment should function as a decision-support tool that assists teacher judgment, encourages student reflection, and strengthens assessment documentation, while still requiring human interpretation, ethical awareness, and contextual sensitivity.

Table 4. Joint Display of Qualitative Validation and Model Interpretation

Data Source	Evidence Found	Convergence with Model Output	Divergence / Deviant Case	Interpretation
Teacher interviews	Teachers reported that high model scores often matched students' discipline, empathy, and participation	Confirms pedagogical relevance	Teachers rejected the use of AI as final judgment	AI is useful as supporting evidence
Student interviews	Students felt feedback encouraged self-reflection	Confirms reflective usefulness	Some students needed teacher explanation to understand scores	Feedback must be dialogical
Classroom observation	High-score students generally showed respectful interaction and active participation	Confirms behavioral validity	Some students behaved positively but wrote weak reflections	Affective data must be interpreted multimodally
Document analysis	Digital records supported model integration, especially in schools with stronger documentation	Confirms institutional readiness	Documentation quality varied across schools	Data infrastructure affects model interpretability
Member checking	Participants confirmed most thematic interpretations	Strengthens credibility	Teachers requested careful wording to avoid overclaiming	AI should remain teacher-mediated
Integrated interpretation	Model outputs aligned with several classroom realities but not all cases	Shows partial validation	Reveals limits of single-source prediction	AI-based assessment is best used as decision support

Ethical and Pedagogical Validation of the Proposed Framework

The integrated phase examined whether the proposed framework was ethically acceptable and pedagogically usable in Islamic Religious Education. The analysis showed that the framework could provide supplementary evidence for affective assessment, but it could not independently determine students' religiosity, sincerity, or moral quality. The model inferred patterns from reflective texts, behavioral logs, and digital interaction records; therefore, its outputs should be understood as probabilistic indicators rather than definitive judgments about students' inner spiritual states. This distinction is essential because affective outcomes in IRE involve moral intention, adab, and spiritual awareness, which cannot be fully captured by computational traces alone. Consequently, the framework was positioned as a teacher-mediated decision-support tool rather than an autonomous evaluator.

Ethical validation revealed three main governance concerns: privacy protection, informed consent, and interpretive accountability. Because the dataset involved minors and religious-affective information, the schools and research team applied institutional permission, parental consent, student assent, anonymization, data minimization, and restricted data access before analysis. However, interview and document evidence showed that the degree of procedural standardization varied across schools. SMA Islamic Village had more organized digital documentation and clearer

internal procedures, while SMAN 24 and SMAN 19 relied more heavily on teacher-managed records. This variation did not invalidate the model, but it demonstrated that AI-based affective assessment requires institutional data governance before being used in regular school assessment practices.

Fairness analysis was conducted to examine whether model performance varied across gender, school context, grade level, digital access, writing proficiency, and participation level. The analysis showed that prediction performance was relatively comparable across gender groups, with a difference of less than three percentage points between male and female students. Larger variation appeared across school contexts and documentation quality, particularly between schools with more systematic digital records and those with teacher-centered assessment archives. Performance was also lower among students with limited digital interaction records and shorter reflective texts. These findings indicate that potential bias did not primarily arise from demographic identity alone, but from unequal data completeness, writing expressiveness, and institutional documentation practices.

The error review further clarified the limits of AI inference. Students with high classroom participation but short reflective writing were sometimes classified lower than teachers expected, while students with strong written reflections but inconsistent classroom engagement occasionally received higher predicted affective scores. This pattern shows that the model is more capable of identifying observable and textual indicators of affective development than judging internal intention or sincerity. Therefore, claims of objectivity must be carefully bounded. The model may improve consistency in organizing evidence, but it cannot replace teacher interpretation, moral dialogue, or direct observation. In this sense, computational assessment provides structured evidence, while pedagogical judgment gives that evidence educational meaning.

Pedagogically, teachers considered the framework useful when it was used to identify students who might need further guidance, reinforcement, or reflective dialogue. The model helped teachers notice patterns in students' reflections, participation, and digital engagement that were difficult to track manually across seven learning cycles. Nevertheless, teachers emphasized that assessment of adab, sincerity, and religious awareness must remain embedded in Islamic pedagogical relationships. AI-generated indicators were therefore interpreted through classroom observation, teacher-student dialogue, and ethical consideration. The framework's pedagogical value lies not in replacing conventional assessment, but in strengthening teachers' ability to combine data-based evidence with value-based educational interpretation.

The final framework integrates four layers: quantitative prediction, qualitative interpretation, ethical governance, and Islamic pedagogical mediation. Quantitative prediction identifies patterns in student data; qualitative interpretation verifies whether the patterns are meaningful in classroom realities; ethical governance protects students from misuse, over-surveillance, and unfair classification; and Islamic pedagogical mediation ensures that assessment remains oriented toward guidance, character formation, and human dignity. This layered structure reconciles the tension between computational measurement and the spiritual nature of affective learning. It shows that AI can support affective assessment only when its outputs are treated as partial, contextual, and revisable evidence.

Overall, the ethical and pedagogical validation indicates that the proposed framework is feasible as a supplementary assessment mechanism, but not as a stand-alone instrument for measuring religious character. Its strongest contribution lies in organizing multimodal evidence, identifying affective patterns, supporting teacher reflection, and improving documentation. Its main

limitation lies in the inability to infer inner intention, sincerity, or spiritual depth without human interpretation. Therefore, the study sets strict boundaries: AI can classify observable indicators and learning traces related to affective development, but it cannot determine the authenticity of students' faith, moral intention, or spiritual transformation. These boundaries are essential for responsible implementation in Islamic Religious Education.

Table 5. Ethical and Pedagogical Validation of the Framework

Aspect	Evidence / Finding	Boundary of AI Inference	Pedagogical Implication
Data privacy	Anonymization, data minimization, restricted access, and password-protected files were applied	AI should not process identifiable religious-affective data without safeguards	Schools need standardized data governance policies
Informed consent	Institutional permission, parental consent, student assent, and voluntary participation were applied	AI assessment must not be imposed as compulsory surveillance	Consent must be transparent and educationally non-punitive
Fairness by gender	Performance difference between male and female groups remained below three percentage points	Gender should not be used as a direct affective predictor	Results require fairness monitoring
Fairness by school context	Variation appeared across schools due to differences in documentation quality	AI may reflect institutional data quality rather than student affect alone	Schools need comparable documentation standards
Digital access	Lower data completeness reduced prediction confidence	AI cannot fairly assess students with insufficient digital traces	Alternative teacher-based evidence is needed
Writing proficiency	Short or less expressive reflections increased misclassification risk	AI cannot equate weak writing with weak affective development	Reflection scores require teacher interpretation
Participation level	High classroom participation sometimes contradicted low textual prediction	AI cannot infer sincerity from one data source	Multimodal and observational validation are required
Islamic value alignment	Teachers emphasized adab, amanah, sincerity, and moral dialogue	AI cannot judge faith, intention, or spiritual authenticity	AI must remain a teacher-mediated decision-support tool
Final framework	Integrates prediction, interpretation, ethics, and pedagogy	AI provides probabilistic indicators, not final moral judgment	Supports responsible, contextual, and human-centered assessment

Discussion

The findings show that deep learning can support the operationalization of selected affective indicators in Islamic Religious Education, but only within carefully defined epistemological limits. CNN, RNN, and transformer-based models were able to classify patterns associated with religiosity, empathy, moral awareness, ethical disposition, and spiritual engagement from reflective texts,

behavioral logs, and digital interaction records. However, these outputs should not be interpreted as direct measurements of students' inner faith, sincerity, or spiritual authenticity. Rather, they represent probabilistic classifications of observable and textual learning traces. This distinction is crucial because affective learning in Islamic education includes theological, moral, and pedagogical dimensions that exceed computational representation (AlZu'bi et al. 2022; Ismail et al. 2025; Taye 2023; Tumin and Septariani 2026).

The stronger performance of the transformer-based model at SMAN 24 Kabupaten Tangerang suggests that reflective texts are valuable for identifying semantic patterns related to moral awareness, empathy, and religious meaning-making. Nevertheless, the discussion must avoid reducing reflective language to inner religiosity. Students may articulate Islamic values convincingly in writing, but such articulation does not automatically prove sincerity, adab, or stable moral character. This finding supports the usefulness of NLP-based analysis for interpreting textual learning evidence, while simultaneously confirming the need for teacher mediation and classroom observation (Ravi and Yuan 2024; Sayyi, Mashuri, et al. 2025). In this sense, transformer models provide interpretive support, not theological certainty (Amirudin, Muzaki, and Nurhayati 2026; Purwaningsih and Yahya 2026).

The RNN results at SMAN 19 Kabupaten Tangerang indicate that affective development is better understood as a temporal and developmental process than as a single assessment event. Behavioral logs collected across seven learning cycles helped identify patterns of participation, consistency, and engagement. This finding challenges conventional affective assessment practices that rely heavily on episodic observation or general teacher impressions. However, temporal consistency should not be equated with moral maturity in an absolute sense. Repeated participation may indicate discipline and engagement, but it still requires interpretation through pedagogical dialogue and ethical reflection. Sequence-based modeling is therefore useful for identifying developmental trajectories, but not sufficient for judging inner intention (El Alami, El Batteoui, and Satori 2026; Rustam and Delia Jurcut 2026).

The multimodal model at SMA Islamic Village produced the highest performance because it integrated textual, behavioral, and interactional data. The ablation results showed that the full multimodal model outperformed single-modality configurations, indicating that affective learning evidence is distributed across different forms of data. This supports studies showing that multimodal systems can improve prediction in emotion recognition and learning analytics (Khan et al. 2026; Wu, Mi, and Gao 2025; Yazici et al. 2026).. However, the higher performance should be interpreted cautiously because it may also reflect stronger documentation practices at SMA Islamic Village. Thus, multimodal superiority in this study indicates improved evidence integration, not universal validity across all Islamic education settings (Fithriyah 2023; S.h et al. 2024; Sholeh et al. 2023).

Pedagogically, the study contributes a model of AI-supported affective assessment that strengthens, rather than replaces, teacher judgment. Teachers can use model outputs to identify students who may require reinforcement, reflective dialogue, or further guidance. This aligns with research on AI-supported personalization when technological tools remain subordinate to educational purposes (Gligorea et al. 2023; Sahrowi, Harianto, and Sayyi 2025). In Islamic education, however, pedagogical personalization must be framed by values such as ikhlas, amanah, adab, and moral responsibility. Therefore, AI-generated affective indicators should be treated as prompts for

teacher-student dialogue, not as final moral labels. The model is pedagogically meaningful only when embedded within Islamic educational relationships (Sari 2025; Sholeh et al. 2023).

Ethically, this study shows that AI-based affective assessment requires concrete governance architecture, not only general ethical principles. The proposed governance structure includes a school-appointed data steward, restricted data access, periodic bias audits, model documentation, and a human-in-the-loop review protocol. Model outputs should be reviewed by IRE teachers before being used for feedback or intervention. Students should also have an appeal mechanism through which they can question or clarify AI-supported feedback without academic penalty. Bias monitoring should be conducted periodically across gender, grade level, school context, writing proficiency, digital access, and participation level. These procedures respond to concerns about privacy, algorithmic bias, and surveillance pedagogy (Ferrara 2024; Haetami 2025; Haris, Sayyi, and Ismail 2026; Khavenson, Yemini, and Sperduti 2026; Nawaz et al. 2025).

The theoretical contribution of this study should be understood as an initial integrative framework rather than a fully established theory. The framework connects four concepts: multimodal learning evidence, computational prediction, pedagogical interpretation, and Islamic ethical governance. Its mechanism operates through a sequence of data collection, supervised classification, qualitative validation, fairness review, and teacher-mediated feedback. Its central assumption is that affective development leaves partial traces in students' reflections, behavior, and learning interactions. Its boundary condition is equally important: the model can classify observable indicators of affective development, but it cannot determine faith authenticity, sincerity, or spiritual transformation. This clarification prevents premature theorization and aligns the contribution with human-centered AI education (Al-Hail et al. 2024; Hays et al. 2025; Khosravi et al. 2022; Vieriu and Petrea 2025).

Taken together, the findings suggest that the proposed deep learning-based framework represents a significant advancement in affective assessment within Islamic Religious Education. The integration of multimodal learning analytics, qualitative validation, and ethical governance demonstrates that affective constructs can be measured more systematically without neglecting their contextual, moral, and spiritual dimensions. The quantitative results confirm the predictive capability of deep learning architectures, while the qualitative findings reinforce their pedagogical relevance and practical applicability. Furthermore, the incorporation of Islamic ethical principles ensures that technological innovation remains aligned with the broader objectives of character formation and holistic human development. Therefore, the study provides preliminary evidence that AI-supported affective assessment can serve as a reliable, ethically grounded, and educationally meaningful approach for strengthening the quality of Islamic Religious Education in the era of digital transformation.

Taken together, the findings provide preliminary empirical evidence that deep learning can support affective assessment in Islamic Religious Education when combined with multimodal data, qualitative validation, ethical governance, and teacher mediation. The study does not claim that AI can objectively measure religiosity or moral character in a complete theological sense. Instead, it shows that AI can help organize learning evidence, identify affective patterns, and support reflective pedagogical decisions. The most important contribution is therefore not technological automation, but the formulation of a bounded, accountable, and pedagogically mediated framework for using AI in value-oriented education. Such a framework is meaningful only when computational inference

remains accountable to human judgment, Islamic ethics, and the educational aim of holistic character formation.

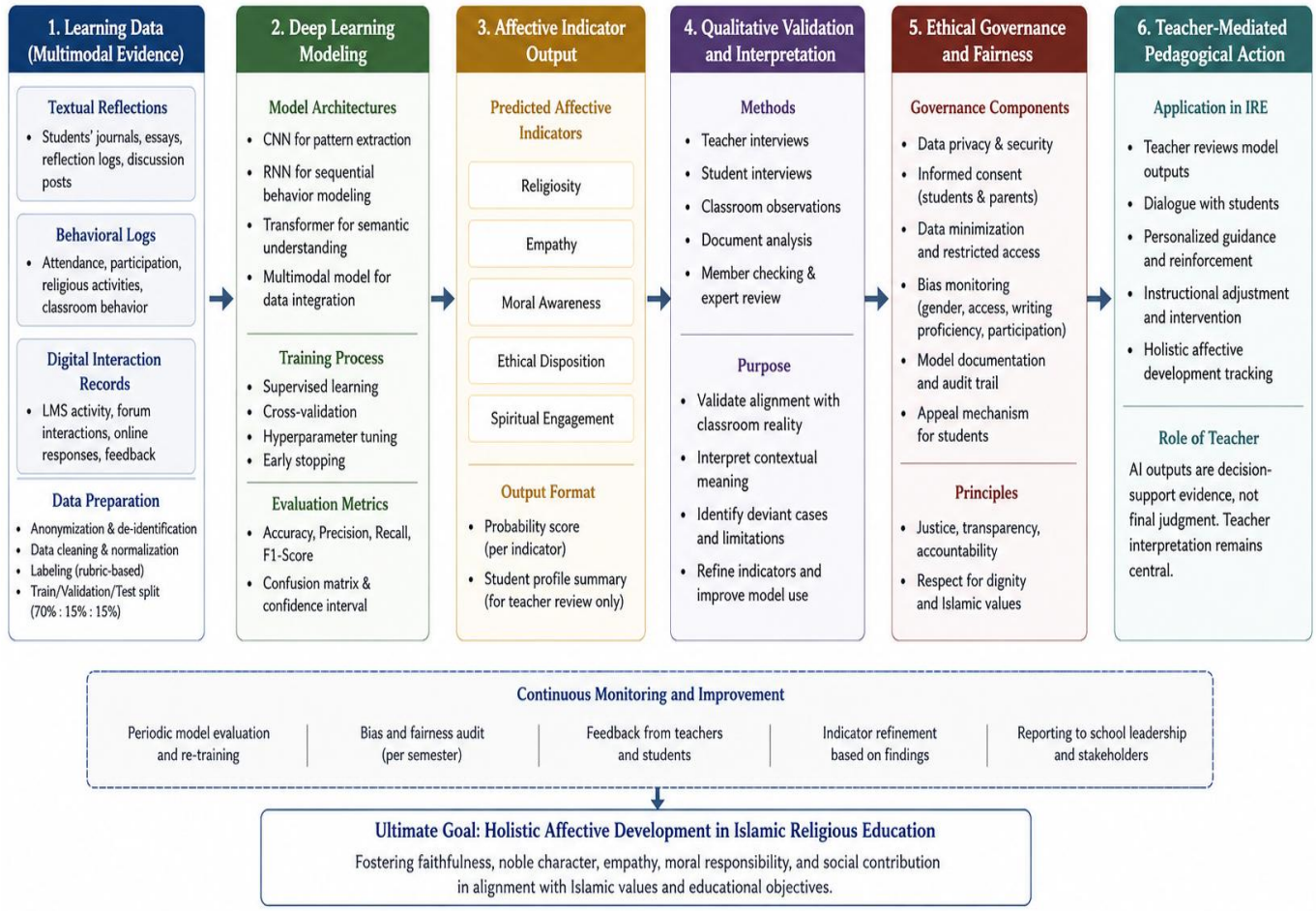

Figure 1. Applied Framework for Deep Learning-Based Affective Assessment in Islamic Religious Education

Figure 1 illustrates an integrated framework for deep learning-based affective assessment in Islamic Religious Education by connecting multimodal learning evidence, computational analysis, qualitative validation, ethical governance, and teacher-mediated pedagogical action. The framework demonstrates that affective assessment should not rely solely on algorithmic prediction but must be strengthened through contextual interpretation, fairness monitoring, and Islamic educational values. By positioning teachers as the primary interpreters of AI-generated outputs, the model promotes a balanced approach that combines technological innovation, ethical responsibility, and holistic character development within contemporary Islamic education.

Conclusion

This study examined the preliminary use of deep learning-based models to support affective assessment in Islamic Religious Education (IRE) at SMAN 24 Kabupaten Tangerang, SMAN 19 Kabupaten Tangerang, and SMA Islamic Village. The findings indicate that CNN, RNN, transformer-based, and multimodal models can classify selected affective indicators, including religiosity, empathy, moral awareness, ethical disposition, and spiritual engagement, based on textual reflections, behavioral logs, and digital interaction records. However, these results should be interpreted cautiously. The models do not directly measure students' inner religiosity, sincerity, adab, or spiritual transformation, but identify observable and textual patterns that may assist teacher

interpretation. Therefore, the proposed framework should be understood as a preliminary, teacher-mediated decision-support model rather than a fully validated standardized instrument for affective assessment.

The study also shows that AI-supported affective assessment requires pedagogical mediation, ethical governance, and contextual validation. Qualitative findings suggest that model outputs may help teachers organize multimodal evidence, identify students who require guidance, and support reflective dialogue, but they cannot replace professional judgment, classroom observation, or moral education. The study is limited to three schools with specific digital documentation practices, a relatively bounded dataset, and preliminary validation procedures; therefore, the findings cannot be generalized to all IRE contexts. Future research should involve larger and more diverse datasets, external validation, longitudinal testing, clearer replication protocols, and rigorous fairness audits across demographic, pedagogical, and institutional variables. Thus, this study contributes an initial empirically informed framework for responsible, ethically bounded, and pedagogically meaningful use of deep learning in Islamic Religious Education.

References

Adewale, M. D., Azeta, A., Abayomi-Alli, A., & Sambo-Magaji, A. (2024). Impact of artificial intelligence adoption on students' academic performance in open and distance learning: A systematic literature review. *Heliyon, 10*(22), Article e40025. https://doi.org/10.1016/j.heliyon.2024.e40025

Al-Attas, S. M. N. (1991). *The concept of education in Islam: A framework for an Islamic philosophy of education*. International Institute of Islamic Thought and Civilization.

Al-Hail, M., Zguir, M. F., & Koç, M. (2024). Exploring digital learning opportunities and challenges in higher education institutes: Stakeholder analysis on the use of social media for effective sustainability of learning–teaching–assessment in a university setting in Qatar. *Sustainability, 16*(15), Article 6413. https://doi.org/10.3390/su16156413

Alismaiel, O. A. (2021). Using structural equation modeling to assess online learning systems' educational sustainability for university students. *Sustainability, 13*(24), Article 13565. https://doi.org/10.3390/su132413565

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data, 8*(1), Article 53. https://doi.org/10.1186/s40537-021-00444-8

AlZu'bi, S., Abu Zitar, R., Hawashin, B., Abu Shanab, S., Zraiqat, A., Mughaid, A., Almotairi, K. H., & Abualigah, L. (2022). A novel deep learning technique for detecting emotional impact in online education. *Electronics, 11*(18), Article 2964. https://doi.org/10.3390/electronics11182964

Amirudin, A., Muzaki, I. A., & Nurhayati, S. (2026). Reflective digital pedagogy in Islamic religious education and religious character formation in higher education. *International Journal of Learning, Teaching and Educational Research, 25*(2), 763–782. https://doi.org/10.26803/ijlter.25.2.34

Ayik, B., Gu, D., Zan, Y., Kim, S., & Kim, W. L. (2026). Human vs. AI: Evaluating thematic analysis with ChatGPT, QInsights, ATLAS.ti AI, and MAXQDA AI Assist. *Qualitative Inquiry*. Advance online publication. https://doi.org/10.1177/10778004251412874

Batista, J., Mesquita, A., & Carnaz, G. (2024). Generative AI and higher education: Trends, challenges, and future directions from a systematic literature review. *Information, 15*(11), Article 676. https://doi.org/10.3390/info15110676

Bloom, B. S., Krathwohl, D. R., & Masia, B. B. (1964). *Taxonomy of educational objectives: The classification of educational goals. Handbook II: Affective domain*. Longman Group.

Costa, J. (2024). Mixed methods in educational large-scale studies: Integrating qualitative perspectives into secondary data analysis. *Education Sciences, 14*(12), Article 1347. https://doi.org/10.3390/educsci14121347

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). SAGE.

Eitel, F., Schulz, M.-A., Seiler, M., Walter, H., & Ritter, K. (2021). Promises and pitfalls of deep neural networks in neuroimaging-based psychiatric research. *Experimental Neurology, 339*, Article 113608. https://doi.org/10.1016/j.expneurol.2021.113608

El Alami, A., El Batteoui, I., & Satori, K. (2026). Real-time Bernoulli-based sequence modeling for efficient intrusion detection in network flow data. *Journal of Cybersecurity and Privacy, 6*(1), Article 32. https://doi.org/10.3390/jcp6010032

Fàbregues, S., Sáinz, M., Romano, M. J., Escalante-Barrios, E. L., Younas, A., & López-Pérez, B.-S. (2023). Use of mixed methods research in intervention studies to increase young people's interest in STEM: A systematic methodological review. *Frontiers in Psychology, 13*, Article 956300. https://doi.org/10.3389/fpsyg.2022.956300

Ferrara, E. (2024). Fairness and bias in artificial intelligence: A brief survey of sources, impacts, and mitigation strategies. *Sci, 6*(1), Article 3. https://doi.org/10.3390/sci6010003

Fithriyah, I. (2023). Installation of religious moderation values in multi ethnic and religious communities in Polagan Galis Pamekasan Village. *FIKROTUNA: Jurnal Pendidikan dan Manajemen Islam, 12*(2), 198–217. https://doi.org/10.32806/jf.v12i02.7350

Fithriyah, I. (2026). The application of spiritual values in a cognitive framework for coping with students' academic stress. *Jurnal Konseling Pendidikan Islam, 7*(2). https://doi.org/10.32806/jkpi.v7i2.1590

Gligorea, I., Cioca, M., Oancea, R., Gorski, A.-T., Gorski, H., & Tudorache, P. (2023). Adaptive learning using artificial intelligence in e-learning: A literature review. *Education Sciences, 13*(12), Article 1216. https://doi.org/10.3390/educsci13121216

Gouseti, A., James, F., Fallin, L., & Burden, K. (2025). The ethics of using AI in K-12 education: A systematic literature review. *Technology, Pedagogy and Education, 34*(2), 161–182. https://doi.org/10.1080/1475939X.2024.2428601

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., & Chen, T. (2018). Recent advances in convolutional neural networks. *Pattern Recognition, 77*, 354–377. https://doi.org/10.1016/j.patcog.2017.10.013

Haetami, H. (2025). AI-driven educational transformation in Indonesia: From learning personalization to institutional management. *Al-Ishlah: Jurnal Pendidikan, 17*(2), 1819–1832. https://doi.org/10.35445/alishlah.v17i2.7448

Halstead, M. (2004). An Islamic concept of education. *Comparative Education, 40*(4), 517–529. https://doi.org/10.1080/0305006042000284510

Haris, A. B., Sayyi, A., & Ismail, I. (2026). Strengthening religious moderation through Islamic religious education learning: Integrative planning, dialogical instruction, and continuous evaluation. *Jurnal Cendekia: Media Komunikasi Penelitian dan Pengembangan Pendidikan Islam, 18*(1), 205–221. https://doi.org/10.37850/cendekia.v18i01.1290

Hassannataj Joloudari, J., Maftoun, M., Nakisa, B., Alizadehsani, R., Yadollahzadeh-Tabari, M., & Gaftandzhieva, S. (2026). Complex emotion recognition system using basic emotions via facial expression, electroencephalogram, and electrocardiogram signals: A review. *Frontiers in Psychology, 17*, Article 1682883. https://doi.org/10.3389/fpsyg.2026.1682883

Hays, L., Miller-Young, J., & McCollum, B. (2025). *Educational technology and the scholarship of teaching and learning: Asking questions about our practices*. TRU Open Press.

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.

Humaida, I. K. W., & Mukmin. (2025). Digital Islamic studies learning methodology for Generation Z. *Chalim Journal of Teaching and Learning, 5*(2), 155–166. https://doi.org/10.31538/cjotl.v5i2.2748

Ifenthaler, D., & Schumacher, C. (2016). Student perceptions of privacy principles for learning analytics. *Educational Technology Research and Development, 64*(5), 923–938. https://doi.org/10.1007/s11423-016-9477-y

Intania, N., Ruwandi, R., & Saerozi, S. (2026). Islamic education in the digital era: An integrative conceptual study of the tech–touch–teach paradigm. *Edukasi Islami: Jurnal Pendidikan Islam, 15*(1), 81–96. https://doi.org/10.30868/ei.v15i01.9627

Ismail, I., Takwil, M., Sayuti, A. F., Ridho, A., Sayyi, A., & Fauzi, A. (2025). The synthesis of turath and modern instructional design: Enhancing fiqh al-mawarith mastery at Maktab Nubdzat Al-Bayan. *EDURELIGIA: Jurnal Pendidikan Agama Islam, 9*(3), 243–257. https://doi.org/10.33650/edureligia.v9i3.11625

Johnson, J. L., Adkins, D., & Chauvin, S. (2020). A review of the quality indicators of rigor in qualitative research. *American Journal of Pharmaceutical Education, 84*(1), Article 7120. https://doi.org/10.5688/ajpe7120

Khan, B., Usman, M., & Binsawad, M. (2026). A novel hybrid model for emotion detection in text through sequential and transformer-based approaches: LSTM enhanced RoBERTa (LER). *Scientific Reports, 16*, Article 2224. https://doi.org/10.1038/s41598-025-31984-1

Khavenson, T., Yemini, M., & Sperduti, V. (2026). Navigating ethical approval challenges: A case of cross-national multi-method social science project. *Methods in Psychology, 14*, Article 100223. https://doi.org/10.1016/j.metip.2025.100223

Khosravi, H., Buckingham Shum, S., Chen, G., Conati, C., Tsai, Y.-S., Kay, J., Knight, S., Martinez-Maldonado, R., Sadiq, S., & Gašević, D. (2022). Explainable artificial intelligence in education. *Computers & Education: Artificial Intelligence, 3*, Article 100074. https://doi.org/10.1016/j.caeai.2022.100074

Li, P., Guo, S., Zhang, Y., Hu, H., Cheng, T., Xu, B., Zeng, K., Huang, T., Dong, Z., Huo, B., Lin, J., Wen, H., & Sun, B. (2026). Multimodal deep learning for inflammatory bowel disease: A new frontier in cellular and molecular biomarker discovery to clinical translation. *Journal of Biological Engineering, 20*, Article 49. https://doi.org/10.1186/s13036-026-00637-w

Lian, H., Lu, C., Li, S., Zhao, Y., Tang, C., & Zong, Y. (2023). A survey of deep learning-based multimodal emotion recognition: Speech, text, and face. *Entropy, 25*(10), Article 1440. https://doi.org/10.3390/e25101440

Liu, Y., Zhang, H., Jiang, M., Chen, J., & Wang, M. (2024). A systematic review of research on emotional artificial intelligence in English language education. *System, 126*, Article 103478. https://doi.org/10.1016/j.system.2024.103478

Mienye, I. D., & Swart, T. G. (2024). A comprehensive review of deep learning: Architectures, recent advances, and applications. *Information, 15*(12), Article 755. https://doi.org/10.3390/info15120755

Mohammadi, M., Tajik, E., Martinez-Maldonado, R., Sadiq, S., Tomaszewski, W., & Khosravi, H. (2025). Artificial intelligence in multimodal learning analytics: A systematic literature review. *Computers & Education: Artificial Intelligence, 8*, Article 100426. https://doi.org/10.1016/j.caeai.2025.100426

Nawaz, M., Awan, N., Ahmed, S., & Mustafa, A. (2025). Surveillance pedagogy: The psychological and pedagogical risks of AI-based behavioral analytics in digital classrooms. *ACADEMIA International Journal for Social Sciences, 4*(3), 1995–2010. https://doi.org/10.63056/ACAD.004.03.0508

Phelps, E. E., Tutton, E., Griffin, X., & Baird, J. (2020). A mixed-methods systematic review of patients' experience of being invited to participate in surgical randomised controlled trials. *Social Science & Medicine, 253*, Article 112961. https://doi.org/10.1016/j.socscimed.2020.112961

Purwaningsih, R., & Yahya, M. S. (2026). The importance of Islamic religious education and moral education in shaping the character of elementary school students. *Jurnal Konseling Pendidikan Islam, 7*(2), 219–231. https://doi.org/10.32806/jkpi.v7i2.1634

Rahmawati, E. M., & Wahyuningsih, R. (2025). Concept of Islamic education according to Buya Hamka. *Al-Afkar, Journal for Islamic Studies, 8*(1), 360–370. https://doi.org/10.31943/afkarjournal.v8i1.1029

Raj, R., & Kos, A. (2025). An extensive study of convolutional neural networks: Applications in computer vision for improved robotics perceptions. *Sensors, 25*(4), Article 1033. https://doi.org/10.3390/s25041033

Ravi, K., & Yuan, J.-S. (2024). Ideological orientation and extremism detection in online social networking sites: A systematic review. *Intelligent Systems with Applications, 24*, Article 200456. https://doi.org/10.1016/j.iswa.2024.200456

Rodríguez-Ortiz, M. Á., Santana-Mancilla, P. C., & Anido-Rifón, L. E. (2025). Machine learning and generative AI in learning analytics for higher education: A systematic review of models, trends, and challenges. *Applied Sciences, 15*(15), Article 8679. https://doi.org/10.3390/app15158679

Rustam, F., & Jurcut, A. D. (2026). A simulation-based unified framework for real-time malicious traffic detection in multi-environment networks. *Computer Networks, 281*, Article 112220. https://doi.org/10.1016/j.comnet.2026.112220

Ryans, D. G., & Krathwohl, D. R. (1965). Stating objectives appropriately for program, for curriculum, and for instructional materials development. *Journal of Teacher Education, 16*(1), 83–92. https://doi.org/10.1177/002248716501600115

Sahrowi, A., Harianto, S., & Sayyi, A. (2025). Implementasi pembelajaran berbasis humanistik dalam meningkatkan spiritualitas santri di Pesantren Darul Ulum Banyuanyar Pamekasan. *Jurnal Studi Pendidikan Agama Islam, 1*(1), 14–24. https://doi.org/10.32806/jspai.v1i1.912

Saputra, R. F. A., Ridha, M., & Sulaimon, J. T. (2025). Deep learning applications in primary education: A systematic literature review of emerging trends, challenges, and opportunities. *Jurnal Pendidikan Progresif, 15*(3), 1785–1810. https://doi.org/10.23960/jpp.v15i3.pp1785-1810

Sari, W. D. (2025). Transforming the Islamic education curriculum for the Society 5.0 era: Integrating technology, ethics, and pedagogy. *Al-Ishlah: Jurnal Pendidikan, 17*(3), 5635–5643. https://doi.org/10.35445/alishlah.v17i3.7186

Sayyi, A., Asmuki, A., Juandi, W., Alimin, M., & Fithriyah, I. (2025). Bridging tradition and multiculturalism in Islamic jurisprudence education. *Scaffolding: Jurnal Pendidikan Islam dan Multikulturalisme, 7*(3), 859–879. https://doi.org/10.37680/scaffolding.v7i3.8002

Sayyi, A., Mashuri, S., Afandi, A., & Alanesi, A. S. H. G. (2025). Moderate Islamic education curriculum design: Realizing tolerance amidst social diversity in the era of Society 5.0. *Jurnal Konseling Pendidikan Islam, 6*(3), 261–277. https://doi.org/10.32806/jkpi.v6i3.1184

Sheik Mohamed, S. H., Sayyi, A., Nirmala, M., & Elango, S. (2024). Transformation of environmental education in pesantren: Integration of Islamic values and STEAM approach. *FIKROTUNA: Jurnal Pendidikan dan Manajemen Islam, 13*(2), 254–268. https://doi.org/10.32806/jf.v14i2.684

Shingjergji, K., Iren, D., Urlings, C., & Klemke, R. (2026). Affective computing in online higher education: A systematic literature review. *Computers & Education: Artificial Intelligence, 10*, Article 100499. https://doi.org/10.1016/j.caeai.2025.100499

Sholeh, M., Rusydiyah, E. F., & Bakar, M. Y. A. (2024). Integration of AI chatbots in Islamic religious education: Potential and challenges from a doctoral student perspective. *Al-Ishlah: Jurnal Pendidikan, 16*(2), 2105–2121. https://doi.org/10.35445/alishlah.v16i2.5409

Sholeh, M. I., Rohman, H., Suwandi, E. A., Akhyak, Efendi, N., & Muhajir, A. (2023). Transformation of Islamic education: A study of changes in the transformation of the education curriculum. *Jurnal Pendidikan Agama Islam, 20*(1), 33–50. https://doi.org/10.14421/jpai.v20i1.6770

Silmi, K. P., Adkins-Jackson, P., Meléndrez, B., Dang, N., Zeray, S., Cardenas-Iniguez, C., Pugh, E., Perez, Y., Cervantes, N., Peters, P., Hammonds, S., Covarrubias Reyna, I., Álvarez, D., Arias, M., Murillo, J., & Gonzalez, M. (2025). Integrating qualitative insights with large secondary data: A protocol for a community-engaged mixed-methods study on adolescent substance use. *Frontiers in Public Health, 13*, Article 1664492. https://doi.org/10.3389/fpubh.2025.1664492

Suter, W. N. (2012). Qualitative data, analysis, and design. In W. N. Suter (Ed.), *Introduction to educational research: A critical thinking approach* (2nd ed., pp. 342–386). SAGE Publications. https://doi.org/10.4135/9781483384443.n12

Sutiah, & Supriyono. (2024). Enhancing online learning quality: A structural equation modeling analysis of educational technology implementation during the COVID-19 pandemic. *Telematics and Informatics Reports, 16*, Article 100175. https://doi.org/10.1016/j.teler.2024.100175

Taye, M. M. (2023). Theoretical understanding of convolutional neural network: Concepts, architectures, applications, future directions. *Computation, 11*(3), Article 52. https://doi.org/10.3390/computation11030052

Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences*. SAGE Publications.

Tumin, & Septariani, S. F. (2026). Exploring the values of the Prophet Muhammad's Hijrah through a deep learning approach in Islamic education at primary school. *Tafkir: Interdisciplinary Journal of Islamic Education, 7*(2), 432–444. https://doi.org/10.31538/tijie.v7i2.2524

Valentijn, P. P., Schepman, S. M., Opheij, W., & Bruijnzeels, M. A. (2013). Understanding integrated care: A comprehensive conceptual framework based on the integrative functions of primary care. *International Journal of Integrated Care, 13*, Article e010. https://doi.org/10.5334/ijic.886

Vărzaru, A. A. (2025). The digital economy and sustainable development goals: A predictive analysis of the interconnection between digitalization and sustainability in EU countries. *Systems, 13*(6), Article 398. https://doi.org/10.3390/systems13060398

Vieriu, A. M., & Petrea, G. (2025). The impact of artificial intelligence (AI) on students' academic development. *Education Sciences, 15*(3), Article 343. https://doi.org/10.3390/educsci15030343

Vistorte, A. O. R., Deroncele-Acosta, A., Ayala, J. L. M., Barrasa, A., López-Granero, C., & Martí-González, M. (2024). Integrating artificial intelligence to assess emotions in learning environments: A systematic

literature review. *Frontiers in Psychology, 15*, Article 1387089. https://doi.org/10.3389/fpsyg.2024.1387089

Wiese, L. J., Patil, I., Schiff, D. S., & Magana, A. J. (2025). AI ethics education: A systematic literature review. *Computers & Education: Artificial Intelligence, 8*, Article 100405. https://doi.org/10.1016/j.caeai.2025.100405

Wu, Y., Mi, Q., & Gao, T. (2025). A comprehensive review of multimodal emotion recognition: Techniques, challenges, and future directions. *Biomimetics, 10*(7), Article 418. https://doi.org/10.3390/biomimetics10070418

Yazici, A., Kucukyilmaz, T., Dokeroglu, T., Sharipbay, A., Lee, M.-H., & Tyler, B. (2026). State-of-the-art multimodal emotion recognition: A comprehensive survey and taxonomy. *Intelligent Systems with Applications, 30*, Article 200642. https://doi.org/10.1016/j.iswa.2026.200642

Zaharah, Z., Basyit, A., Husein, M. T., Fauzi, A., Arif, Z., & Sina, I. (2024). Revolutionizing learning: The impact of artificial intelligence on Islamic education and the wave of transformation. *Al-Ishlah: Jurnal Pendidikan, 16*(4), 5685–5697. https://doi.org/10.35445/alishlah.v16i4.6078

Zhang, H., Liu, Y., Jiang, M., Chen, J., Wang, M., & Paas, F. (2025). Emotional artificial intelligence in education: A systematic review and meta-analysis. *Educational Psychology Review, 37*(4), Article 106. https://doi.org/10.1007/s10648-025-10086-4

Zurc, J., & Laaksonen, C. (2023). Effectiveness of health promotion interventions in primary schools—A mixed methods literature review. *Healthcare, 11*(13), Article 1817. https://doi.org/10.3390/healthcare11131817