Deep Learning Framework For Accurate Quranic Speech Recognition

Khalid Almeman kmeman@qu.edu.sa
Unit of Scientific Research, Applied College, Qassim University,
Qassim, Saudi Arabia

Corresponding Author: Khalid Almeman

Abstract

The paper suggests a deep learning-based architecture of Quranic speech recognition with specific aim to address prosodic and phonetic demands of Quranic recitation. The model entails Tajweed conscious linguistic and phonetic articulation, studying and deduction, evaluation and feedback, ethical and religious supervision and Quranic information. To facilitate benchmarking and reproducible system development in future, it outlines operational pipeline and quantifiable artifacts (orthographic, phonetic and Tajweed outputs). The framework is installed in a way to enable the assessment of education, accessibility and the devoted transmission of Quranic recitation by viewing Tajweed as inner phonological restraint.

Keywords: Deep learning, Neural networks, Phonetics, Speech recognition, Quranic recitation.

1. INTRODUCTION

The Holy Quran is of spiritual importance and its recitation is aided by phonetic and melodic principles of Tajweed, and other recitations referred to as Qira'at [1, 2]. These features, which distinguish Quranic Arabic from Modern Standard Arabic and pose challenges to automatic speech recognition (ASR) systems, include pharyngealized and emphatic sounds [3, 4], unusual phonetic phenomena [4], and regular rhythmic patterns [3]. These classic Arabic structures find it difficult to deal with such variations, especially in educational or religious settings where accuracy and precision are highly regarded and the results are low accuracy and misrecognition [5–7]. Moreover, regional accents and differences in Quranic recitation (Qira'at) are another difficulty for ASR systems [8, 9].

Despite the fact that over time, the general areas of deep learning methods have evolved [10, 11], their use in the Quranic recitation is still in its infancy [12, 13]. Moreover, their performance is subject to the quality of available audio and textual data that must be deciphered with the appropriate phonetic information and diacritic signs to guarantee the accuracy of the system [14, 15]. Such data can be useful in training models to identify the peculiarities of Quranic recitations, such as intonation and phonetic variations.

Citation: Khalid Almeman. Deep Learning Framework For Accurate Quranic Speech Recognition. **Advances in Artificial Intelligence and Machine Learning**. 2026. (Ahead of Print). https://dx.doi.org/10.54364/AAIML.2026.62291

The development of neural networks, long short-term memory (LSTM) and transformer-based models [16, 17]. has resulted in speech recognition technologies that are developed very rapidly. The most recent developments in transformer-based systems, such as Wav2Vec 2.0 [18] and attention-based decoders [19] have achieved the highest performance in ASR which is why they are used in Quranic contexts. For example, an end-to-end model for Quranic ASR achieves a word error rate at score of 6.16% and a character error rate of 1.98%. That is above the current Convolutional Neural Network-Recurrent Neural Network (CNN-RNN) models [20].

In addition to ASR, researchers have focused more on the identification and correction of Tajweed rule violations. Al Harere and Al Jallad [14] developed an LSTM-based model using Mel-frequency cepstral coefficients (MFCC) features that are trained on the QDAT dataset. The model could easily detect Tajweed mispronunciations with more than 95% accuracy across all the regulations, such as Idghaam and Ikhfa. Omran et al. [21] employed CNNs to specify the Qalqala rule at the letter level. They achieved promising results. Building on this work, Al Harere and Al Jallad introduce a CNN-bidirectional GRU framework that is based on the phoneme level for upscaled rule detection [22].

Phoneme-level analysis has also brought reciter identification and speaker verification into the spotlight. A case in point is the application of a CNN trained on MFCC representations, which reached an impressive accuracy level of 99.66% in the identification of different Quranic reciters [23, 24]. NASNetLarge transfer learning was used on spectrogram features for distinguishing 20 different reciters and claimed an accuracy of 98.5%. In addition, a recent trial with Mozilla's DeepSpeech showed a gender bias in ASR systems. Male-only reciting trained models could not convey to female voices, thus reinforcing the need for speaker diversity in the construction of datasets [25, 26].

Notwithstanding this progress, a number of limitations remain. The foremost limitation is the application of transformer models specifically customized for Quranic ASR, particularly those that incorporate the modeling process of Tajweed rules. The other limitation is that the majority of the present studies are based on private or limited datasets, which consequently makes it hardly possible to measure or compare the results obtained through the different methods applied. The third limitation is that even though there are systems for Tajweed detection, only a few models have actually carried out the enforcement or evaluation of compliance with Tajweed during the recitation process.

Considering these aspects, this paper presents a Quranic speech recognition system that incorporates Tajweed regulatory units in the modeling system to facilitate automatic checking of the recited pronunciation. The framework describes (i) a dataset design view, which considers Qiraat diversity and speaker demographics, (ii) Tajweed-conscious learning and inference modules which relate acoustic representation to phoneme-level and orthographic outputs, and (iii) an evaluation and feedback layer, which facilitates educational evaluation under ethical and religious oversight. In this way, it demonstrates necessity of specific Quranic recitation tools and places ML-based systems as the auxiliary tools helping learners and teachers to evaluate and improve the accuracy of recitation.

This specification is a faith-conscious and Tajweed-conscious framework that is implementation-ready and is presented in this manuscript. The deliverables of the current work are (i) an operational pipeline, which defines particular AI modules and information flow; (ii) a layered conceptual architecture, which is based on the Quranic linguistic authenticity and ethical-religious supervision; and

(iii) quantifiable output artifacts (orthographic, phonetic and Tajweed compliance maps) which can be used in future benchmarking. The specification suggested is deliberately placed as future work so that it can provide the ability to construct corpus at large scales, train models and evaluate them quantitatively.

2. FRAMEWORK DESIGN

The given framework develops a theoretical architecture of a Quranic speech recognition system which exploits deep learning technologies. The framework defines the fundamental structural, methodological and ethical principles that must be used to design and evaluate any Quranic speech recognition model. The primary aim is to give a unified theoretical framework, which implicitly and explicitly expresses authenticity of linguistic features, high-order skills of machine learning, and holiness of the Quran.

2.1 Conceptual Overview

The framework is divided into five layers. They are interconnected and arranged in a hierarchy where the Data and Corpus Resources are at the bottom, then Linguistic and Phonetic Representation, Learning and Inference, Evaluation and Feedback and finally Ethical and Religious Supervision (FIGURE 1). The upward flow indicates the systematic knowledge extraction of unprocessed data and the continuous integration of the linguistic constraints, learning mechanisms and religiously based supervision.

2.2 Data and Corpus Collecting and Analyzing Layer

The fundamental elements of framework will be the foundation of this layer, specifically, the Quranic Speech Corpora that serves as conceptual framework of the organization of the structured audio-text data.

Corpora are conceptualized as multifaceted structures of knowledge characterized by three major dimensions:

- Phonetic Diversity: This focuses on incorporation of the multiplicity of recitation, including Tajweed, Tarteel and Hada and the simultaneous adoption of regionally based forms of pronunciation.
- Demographic Variability: This factor indicates that there is a large disparity of heterogeneous demographic groups represented. It means that it involves a broad age group, sex and a far more diverse language environment, which brings the required heterogeneity of the population.
- Annotation Schema: This component shows a hypothetical tagging scheme in a way that it fits the audio records with respect to a specific phonetic characteristic, Tajweed rules and morphological characteristics. Thereby it allows stricter analytical interpretation.

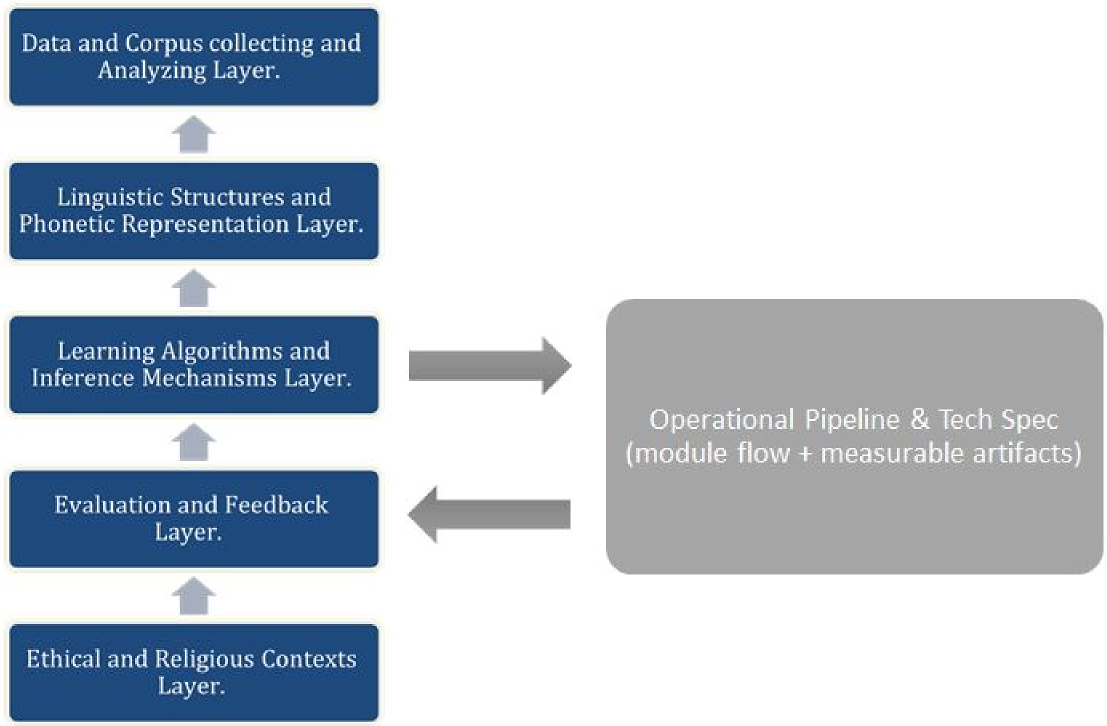

Figure 1: Hierarchical Theoretical Framework

2.2.1 Dataset design blueprint: Non-empirical specification

In order to enhance reproducibility on the design level, the Data and Corpus Layer can be defined using a minimal dataset blueprint. The unit of recording must first be specified (e.g., verse-level segments with time boundaries, optional word- and phoneme-level boundaries to analyze finer details). Second, the text must be standardized as canonical Quranic verse text with complete diacritics, and with an explicit Qira'at label where necessary. Third, the annotation layers must contain (i) phoneme-level transcription with diacritics, (ii) Tajweed rule tags as span-based labels (e.g., Idghaam, Ikhfa, Madd categories) to facilitate verification, and (iii) optional prosodic marks where applicable to rhythmic and lengthening phenomena. Finally, a metadata schema should capture recitation style (e.g., Tarteel/Hadar), speaker demographics, recording conditions and regional background. A standard train/validation/test split protocol should preserve coverage of Qira'at variations and demographic balance to reduce speaker and recitation bias, while maintaining strict textual integrity.

The main concerns of this layer are the data quality and completeness. Instead of focusing on data gathering procedures, it uses theoretical standards that are the basis of the validation of datasets.

2.3 Linguistic Structures and Phonetic Representation Layer

This layer investigates the linguistic basis of Quranic recitation and assumes that Tajweed is not an external set of patterns, but a subset of phonological grammar. A Tajweed Ontology formulates a new approach to relate the recitative properties, e.g., Ikhfa and Idgham, and the relevant phonetic features. Theoretically, a mapping function describes it the following way:

$$\Phi_{Tajweed} : Q_{text} \rightarrow Q_{phoneme}$$

Where:

- Q_{text} refers to the sequence of the Quran verses.
- $Q_{phoneme}$ signifies the abstract set of phonemic articulations that align with Tajweed guidelines.
- $\Phi_{Tajweed}$ consolidates the rules governing the transformation between textual representation and phonetic articulation.

This theoretical model represents the network and defines the interaction between the linguistic and acoustic approaches. It serves as a precise abstraction that educates and guides the process of choosing the model architecture.

2.4 Learning Algorithms and Inference Mechanisms Layer

The assumption regarding the proposed framework is that the process of acquisition is highly abstracted. It takes place through sequential stages of replacing an acoustic representation by a phonological, and finally, an orthographic representation. This change is clearly evident in empirical observation and has measurable effects on performance. It is a transformation in the linguistic competence of the reciter and is accompanied by a series of interconnected phases which link different domains of knowledge:

$$f : A \rightarrow P \rightarrow S$$

Where:

- $\mathcal{A}$ vital component is an acoustic representation, which entails a waveform representation and spectral attributes.
- $\mathcal{P}$ Phonetic (Tajweed -conscious phonemes)
- $\mathcal{S}$ Semantic-textual (Surahs/verses of the Quran)

All transformation processes are based on the theoretical learning operators of the respective type that may include deep models, attention mechanisms, or even mappings of natural predictability. We can formally define these operands as follows:

$$f(\mathcal{A}) = \arg\max_{y \in \mathcal{S}} P(y|\mathcal{A}, \Theta)$$

The symbols, in the current case, are referring to the parameters of the model, which are understood to refer to the relationship between acoustic properties and linguistic properties of the Quran.

The Learning and Inference Layer is defined by a continuous transfer of knowledge in the acoustic, phonetic and semantic domains. The interaction, as shown in FIGURE 2, is not a one-way. The feedback funnels allow the refinement to be iterated so that the representational accuracy of the system can gradually increase. This is a research method that encourages lifelong growth.

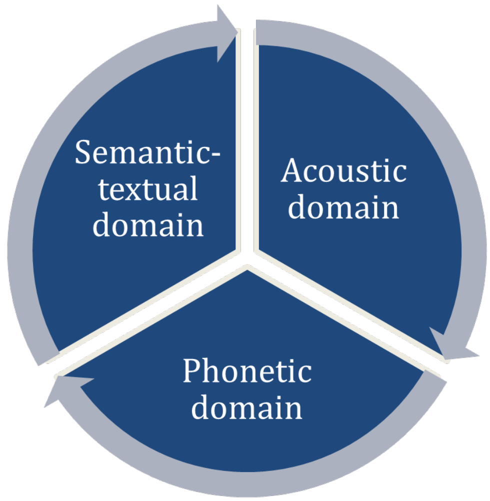

Figure 2: Mutual Exchange in the Knowledge Domains

The symbols in this model are the model parameters, which depict a perfect coordination of the acoustic features and the linguistic form of the Holy Quran. The Learning and Inference Layer is defined by active and ongoing knowledge exchange on the acoustic, phonetic and semantic levels (FIGURE 2). This is two-way process, which thus allows feedback-based refinement to occur, such that the accuracy of the representation is refined over time. In addition, the framework is architecture-neutral: any compatible deep learning paradigm, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs) or transformer-based models, can be incorporated into mapping process, as long as it can be used within the limits of the theoretical framework and fulfills its purpose.

2.5 Operational Pipeline and Technical Specification Layer

Although the given framework is mostly theoretical, it can be converted into an operationally-defined system pipeline that would clarify the meaning of the given framework in regard to actual AI elements, flow of information and quantifiable results. This layer thus acts as an interface between the abstract objectives (linguistic authenticity, Tajweed compliance, ethical supervision) and system modules that can be implemented (acoustic modeling, decoding, phoneme alignment and rule verification). This is not aimed at claiming the deployment of a fully implemented application but defining implementation-ready blueprint that renders the framework reproducible in design and auditable in evaluation.

2.5.1 End-to-end system pipeline: Conceptual but operationally defined

Proposed Quranic ASR pipeline can be described as a series of interrelated modules with evident inputs and outputs, in which Tajweed is not a secondary factor:

(1) Input Acquisition and Gating

- **Input:** Quranic recitation audio signal $\mathcal{A}$ with metadata (Qira'at mode, recitation style, demographic of a speaker).
- **Output:** Authenticated audio segment stream, which is synchronized to reasonable usage terms (education/research) and which complies with the Usage Context Principle.

(2) Signal Standardization and Representation.

- **Input:** Raw waveform $\mathcal{A}$.
- **Processing:** noise handling, segmentation, and feature representation (e.g., log-mel spectrograms) or direct waveform implanting (self-supervised front-ends).
- **Output:** audio embeddings $\mathbf{z}$ which preserve phonetic information of significance in Quranic pronunciation (lengthening, nasalization, emphatics).

(3) Acoustic-to-Phonetic Modeling: Tajweed-Conscious Phoneme Space.

- **Input:** acoustic embeddings $\mathbf{z}$.
- **Processing:** a deep encoder produces a phonetic posterior over a Quranic phoneme inventory consistent with diacritics and classical phonology.
- **Output:** candidates for phoneme sequence $\hat{Q}_{phoneme}$ and confidence profiles which explicitly allow phoneme-level reasoning as well as word-level transcription.

(4) Decoding to Orthographic Quranic Text

- **Input:** $\hat{Q}_{phoneme}$ and a constrained Quranic language model.
- **Processing:** decoding mechanism (CTC/attention or hybrid) maps phonetic hypotheses into Quranic orthography with diacritics as a first-class representation.
- **Output:** predicted verse-level or token-level transcription $\hat{Q}_{text}$, which alignment scores accompany.

(5) Tajweed Rule Verification and Constraint Checking

- **Input:** aligned pairs $\left(Q_{text}, \hat{Q}_{phoneme}\right)$ and the Tajweed ontology.
- **Processing:** rule-level verification identifies expected rule regions (e.g., Idghaam, Ikhfa, categories of Madd) and tests acoustic-phonetic data for conformity.
- **Output:** Compliance of Tajweed map $\hat{R}_{Tajweed}$ containing rule labels, boundary spans and violation indicators.

(6) Feedback and Explainability Interface: Conceptual Visualization

- **Input:** $\hat{Q}_{text}$, $\hat{R}_{Tajweed}$, confidence metrics.
- **Output:** learner-facing feedback that is non-offensive, educationally guided, and aligned with the Transparency Principle (e.g., "rule expected here" rather than theologically loaded language).

(7) Ethical and Religious Supervision Boundary: Overarching Constraint

The output is restricted at both levels due to (i) textual integrity, (ii) restricted contexts of use and (iii) transparency. In this regard, supervision is a continuous gating process which restricts behavior of a system.

This level of operation clearly draws out the difference between:

- **Ideal aspirations:** piousness, reality, religious design.
- **Operational units:** encoder, decoder, aligner, rule checker, feedback module.
- **Measurable results:** accuracy of transcription, consistency of phoneme alignment, detection of Tajweed rule.

2.5.2 Illustrative architecture compatibility: Non-prescriptive, ready for implementation-

- Framework is architecture-neutral, although may be technically pertinent when mapped onto popular ASR paradigms. The following empirical statement can be made:
- **Front-End / Encoder:** Transformer-based acoustic encoders (e.g., wav2vec-style representations) or Conformer-like encoders to encode local phonetic content and global sequence information.
- **Decoding Head:**
  - **CTC-based decoding** for stable alignment properties at the phoneme level; and/or
  - **Attention-based decoding** for richer context modeling in verse-level transcription.
- **Language Model Constraint:** A constraint of the Quranic text that puts a premium on canonical verse structure and diacritically coherent forms, thereby minimizing free-form hallucination and enhancing textual integrity.

Conceptual representation of the proposed framework in FIGURE 3, shows how the framework can be instantiated in a non-prescriptive way without losing Tajweed-conscious verification. The recitation signal is mapped by a shared acoustic encoder backbone (transformer/conformer) into latent representations which diverge into two outputs. The former branch does decode (CTC/attention) with a constraint of a Quranic text to generate the diacritized orthographic transcription Q text. The second branch does phoneme alignment/ mapping and uses a Tajweed verification head with span-based rule tags to generate the compliance map R Tajweed. These two quantifiable artifacts can be directly related to the operational outputs of the framework and can be evaluated using the suggested metrics (LA, PF, SA, PA).

- **Tajweed Layer (Head of Rule Verification):**

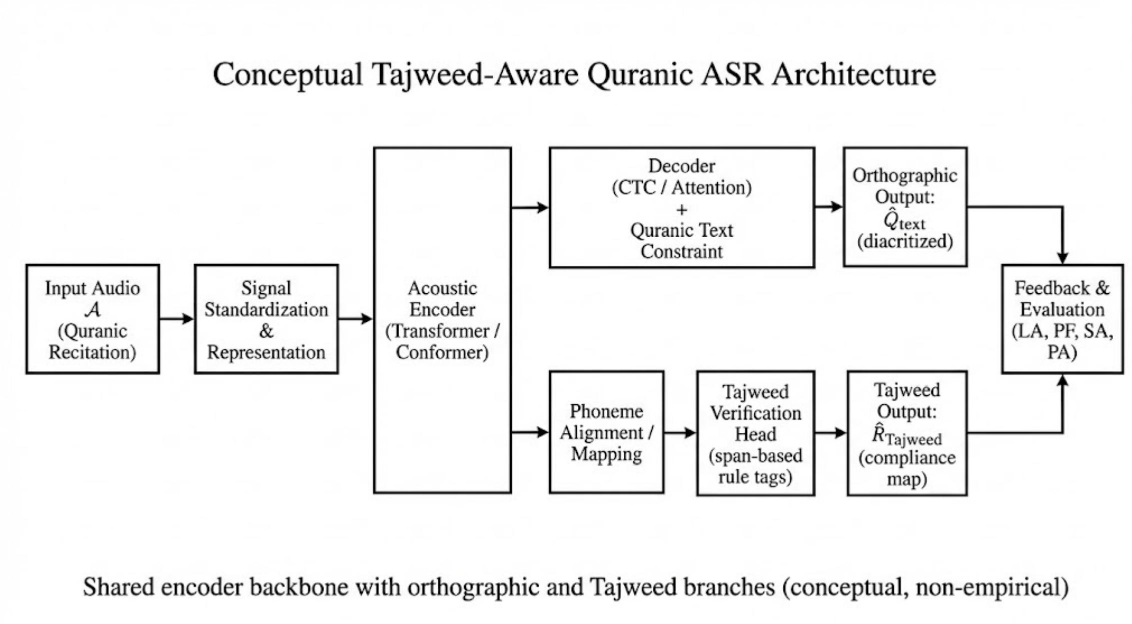

Figure 3: Theoretical Tajweed-conscious Quranic ASR architecture with common encoder backbone and two orthographic decoding and Tajweed verification branches (non-empirical).

A special verification head, which takes phoneme-aligned representations and produces rule-level compliance signals. This head may be conceptualized as a classifier across Tajweed categories with span detection (rule boundary identification). This allows fine-grained educational feedback.

Within the framework spirit, the layer of Tajweed is regarded as official constraint. It influences phonetic interpretation and promotes the fundamental goal of spiritually faithful pronunciation.

2.5.3 Measurable artifacts as operational outputs

In order to prevent the confusion of philosophical vision and engineering deliverables, the pipeline produces three different classes of artifacts:

1. **Orthographic Artifact:** $\hat{Q}_{text}$ (verse/token transcription with diacritics).
2. **Phonetic Artifact:** $\hat{Q}_{phoneme}$ (phoneme sequence and alignment confidence).
3. **Tajweed Artifact:** $\hat{R}_{Tajweed}$ (rule compliance map with span-level interpretation).

These artifacts are specifically made to facilitate subsequent empirical validation and benchmarking. At the same time, they are consistent with the conceptual orientation of the current study.

2.5.4 Worked example: Conceptual tajweed-aware verification

To explain operational meaning of the proposed pipeline, a conceptual walk through is outlined without making any assumption of deployed application or trained model. For instance, a short passage in the Q_{text} with complete diacritics. The system proceeds as follows:

1. **Text-to-Expectation: Tajweed Ontology Projection**
   From Q_{text}, Tajweed ontology projects expected rule regions (e.g., whether nasalization pattern is expected, is Idghaam context present and is a Madd lengthening category conditioned by the orthographic context).
2. **Acoustic-to-Phoneme Hypothesis:**
   With a recitation audio segment, the encoder-decoder generates a phoneme hypothesis $\hat{Q}_{phoneme}$ with time alignment. It is to retain phonetic evidence that is required to confirm Tajweed.
3. **Rule-Level Consistency Check:**
   The verifier compares the phonetic evidence that is in accordance with the anticipated rule areas. The outcome is the compliance map which can be pedagogically communicated.

A simplified presentation is provided in Table 1.

Table 1: Conceptual Tajweed-Aware Verification Output

Stage	Input	Process	Output
Ontology Projection	Q_{text}(diacritics)	Detect expected Tajweed contexts	expected rule spans R_{exp}
Phoneme Inference	audio $\mathcal{A}$	encoder → phoneme posterior	$\hat{Q}_{phoneme}$+ confidence
Alignment	$Q_{text}, \hat{Q}_{phoneme}$	forced/CTC alignment concept	token–phoneme span mapping
Verification	R_{exp}+ aligned phonemes	rule-checking decision	$\hat{R}_{Tajweed}$: compliant/ flagged
Feedback	$\hat{R}_{Tajweed}$	educational rendering	learner-oriented guidance

Provided example shows that proposed framework indicates how restrictions of sacred texts (diacritics and Tajweed rules) can be transformed into working checkpoints that produce measurable artifacts without exceeding the limits of ethical control.

2.6 Evaluation and Feedback Layer

Assessment and Feedback layer explains high conceptual performance criteria which are expected to unify linguistic accuracy with adherence to Quranic principles. It differentiates measure of technical correctness (the precision of speech-to-text transcription) and the measure of spiritual fidelity (observance of the rules of Tajweed and the holiness of Quranic pronunciation) in very obvious way. Theoretical standards are suggested (as presented in Table 2).

Table 2: Measures of evaluation of the deep learning framework of Quranic speech recognition

Metric	Conceptual Description	Interpretation Target
Linguistic Accuracy (LA)	Extent of lexical validity and uniformity	Evaluation of the exactness and specificity of text within a given context.
Phonetic Fidelity (PF)	Adherence to the Proper Pronunciation of Tajweed Sounds	Evaluation of the structural soundness of articulation
Semantic Alignment (SA)	Ensure accurate correspondence between verses and their contextual framework	Assessing the precision of context within the content
Perceived Authenticity (PA)	Assessment relying on personal human perception and opinion	Assessments of perceived fidelity to Quranic teachings

The proposed model proposes involvement of human assessors including reciters and linguists, in model assessment. The assessment establishes a self-perpetuating cycle of semi-mechanized approval through use of professionals as agents with positive feedback. The conceptual interpretation of this integration is as follows:

$$E_{total} = \alpha(LA) + \beta(PF) + \gamma(SA) + \delta(PA)$$

Where α, β, γ and δ are the weighing factors displaying the relation value of every measure.

2.6.1 Example assessment procedure

To elaborate on the manner in which the evaluation may be done in a working manner, a step by step assessment process is provided. (1) WER/CER between the canonical diacritized reference and the predicted transcription $\hat{Q}_{text}$ is used to calculate the technical correctness. 2) Phonetic fidelity is defined as stability of phoneme-level correspondence between Q_{text} and $\hat{Q}_{phoneme}$, which is consistency of acoustic-to-phonetic mapping with Tajweed-aware representation. (3) Rule-wise precision/recall (or F1) on predicted compliance map $\hat{R}_{Tajweed}$, against predicted rule spans based on expected rule spans based on projection of Tajweed ontology on Q_{text}. (4) Obtained indicators are added up to form composite score E_{total} with programmed weight $\alpha, \beta, \gamma, \delta$. In addition to this, borderline cases may be audited using human assessors (qualified reciters and linguists). This assists in assessing the perceived authenticity and in justification of interpretability of system feedback.

2.7 Ethical and Religious Contexts Layer

Moral layer is the least and simultaneously the most in terms of the purity and the domain permitted by the structure. It reaches the point of the rule of veneration in such a manner that the sacredness of the Quran could not be contaminated by machines, even in case they were present at all times. The notions that are necessary and fundamental are:

- **Textual Integrity Principle**: Prohibits any alteration, textual or semantic, to original verses of the Quran or their abbreviation or compilation.
- **Usage Context Principle:** The outputs of the system are strictly to be used in educational and research validation.
- **Transparency Principle:** The decisions made in the model will be in a way that is simple to comprehend and therefore can be justified by scholars. The aforementioned principles set a large framework that controls and directs the conduct and application of any derived system.

The Ethical and Religious Supervision Layer is a regulatory layer that covers all the elements of the framework.

FIGURE 4 sequence shows the significance of this layer in the integrity of morality-based systems. This enabled intellectual and linguistic interactions without going beyond their boundaries. The drawing itself is representation of combination of technological thinking and religious rule. They co-exist and help each another.

Central symbol of the model is the circle, which, simultaneously, symbolizes the ethical and religious limits that offer control over the other components: the Learning, Evaluation, Linguistic and Data layers. The arrows towards the center are used to show the controlled interaction between these components which emphasizes the rule that any religious practice is subject to this layer.

2.8 Limited Interface and Conceptual Visualization

As a way of proving theoretical feasibility, the network has a conceptual user interface which is meant to be illustrative tool and not to be deployed in practice. This interface displays such processes as phoneme detection, Tajweed mapping and feedback flow. The key highlights are:

- The figure provides graphical representation of the signal, which has determined the recitations.
- The phonetic components that coincide with the Tajweed annotations are given.
- Indications of conceptual mistakes are brought up, which justify theoretical disparities between the real and ideal pronunciations.

In this description of abstraction studies, there is an assumption that there is a cognitive-based interface between human awareness and computational processes.

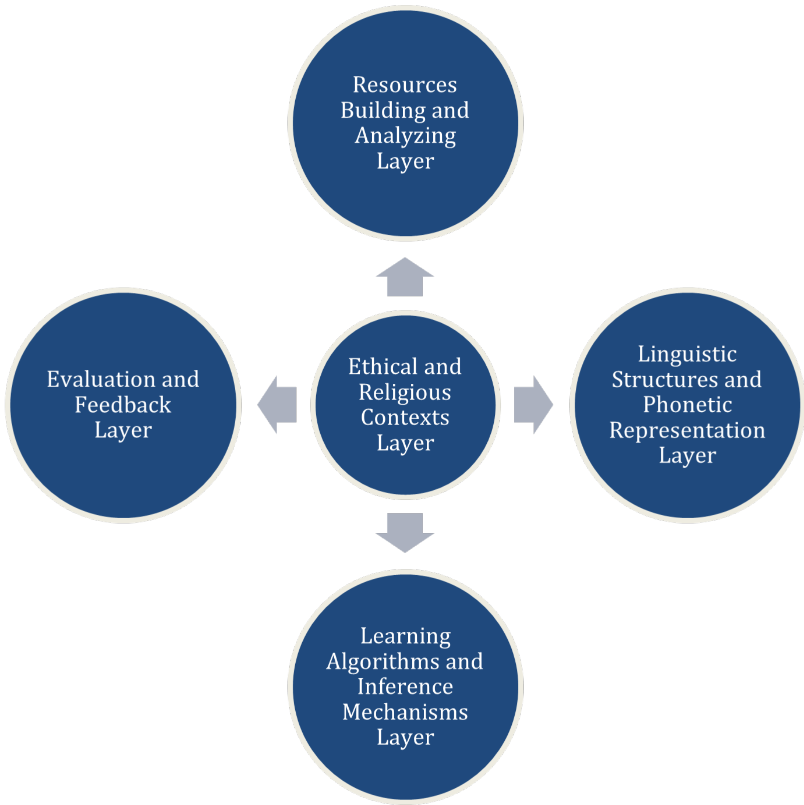

Figure 4: Supervisory Layer in Ethical and Religious Contexts

2.9 Synthesis of the Framework

TABLE 3 provides brief philosophical overview of the framework, which is accurate enough in computation and respectful of religious perspectives.

Table 3: Exemplary Cross-Layer Interface Interaction Matrix

Layer	**Inputs**	**Outputs**	**Supervising Principle**
Data and Corpus Collecting and Analyzing Layer	Quranic corpora, metadata	Annotated phoneme sequences	Completeness
Linguistic Structures and Phonetic Representation Layer	Texts units	Phoneme ontology	Phonetic Authenticity
Learning Algorithms and Inference Mechanisms Layer	Acoustic information	Probabilistic mappings	Optimization
Evaluation and Feedback Layer	Predictions	Composite score	Balance & Validity
Ethical and Religious Contexts Layer	All components	Compliance feedback	Reverence

3. DISCUSSION

The article introduces Quranic speech recognition as a religiously-oriented and linguistically-focused issue. Based on this, the proposed layers incorporate phonetic authenticity, Tajweed ontology, learning and inference, evaluation and ethical-religious supervision to support technical correctness and spiritually faithful recitation assessment.

3.1 Theoretical Implications

In essence, the framework is multi-layered structure of epistemic modalities. It is founded on fact that one can find the linguistic elements of the Quran that cannot be attributed to statistical efficacy, and combine them with linguistic, phonological, semantic and ethical elements on a comprehensive level. Theoretical balance on these aspects can be represented symbolically as:

$$E_{total} = \alpha(LA) + \beta(PF) + \gamma(SA) + \delta(PA)$$

Where:

- LA = Linguistic Accuracy,
- PF = Phoonetic Fidelity,
- SA = Semantic Alignment,
- PA = Perceived Authenticity,

The coefficients show the significance of specific factors in comparison with the overall evaluation of the system.

The formula is explanatory; it determines the view that the effectiveness of Quranic AI technology is to be measured in terms of moral and linguistic values. In addition, the proposed architecture sets ethical standard within the learning environment where all the algorithmic processes must follow assumed moral restrictions to the latter. Theoretically, this relationship between the learning processes and ethics can be represented as:

$$L^* = f(A, P, S) \ subject \ to \ R(E) \leq \tau$$

Here:

- L^* is the maximum level of education performance, which is assessed using different indicators.
- Acoustic, phonetic and semantic categories are represented by *A, P,* and *S* respectively.
- *R(E)* examines the potential ethical consequences and establishes the highest acceptable threshold of practice of religious obligations.

This metaphorical depiction shows idea of restricted learning where AI can be given the opportunity to develop only to the extent of ethical standards.

3.2 Comparative Analysis

To situate theoretical development appropriately, TABLE 4 requires a conventional speech recognition module and the proposed Quranic framework.

Table 4: A Detailed Comparative Analysis of Theoretical Dimensions in Conventional Versus Quranic Frameworks

Dimension	Conventional ASR Models	Proposed Quranic Framework
Orientation	Technical; accuracy-driven	Holistic; value-driven
Data Representation	Generic Arabic corpora	Quranic corpora with Tajweed rules and Qira'at variations
Learning Cycle	Linear: Train → Test	Circular: Learn ↔ Evaluate ↔ Refine
Context Awareness	Phonetic and lexical	Phonetic, semantic, and religious context
Ethical Supervision	External or absent	Integral supervisory layer
Evaluation Metrics	Word Error Rate (WER) CER (Character Error Rate)	LA, PF, SA, PA composite
Human Involvement	Minimal	Central (reciter, linguist, scholar, specialists)
Knowledge View	Data-driven	Knowledge-centered

This comparison shows a change in the mindset, moving away from the classic focus on "engineering accuracy" to "epistemic integrity" being the most important attribute. Traditional systems use error rates have to characterize ASR models as a means of increasing precision. However, the proposed model focuses on linguistic accuracy and theologically appropriateness, ethical soundness and semantic integrity. It goes beyond mechanical correctness to multidimensional correctness based on meaning and reverence.

3.3 Integration of Knowledge Domains

Theoretical progress in the new system was one of the major factors that led to the integration of the separate areas of knowledge into single epistemology that was coherent and easy to understand. The Quran recitation has scientific structure and a spiritual intent. This is why the five layers of the framework merge into a logical entity.

The model is based om the Linguistic & Phonetic Layer. This Layer is the system of pronunciation of Arabic in Quranic recitation and reflects phonetic characteristics of Tajweed. The process from the Linguistic & Phonetic Layer to the Learning Layer makes the linguistic features more abstract and then transforms them into cognitive representations. The Evaluation Layer, thus, also plays a role in completing the feedback loop. It interprets the technical results into different meanings, which makes it possible for the continuous process of refinement and improvement to occur. This repetitive, cyclical operation does not merely produce procedural operations. It is through this kind of philosophical mode that knowledge is enriched and enhanced, through the continuous process of reflection, regression and rectification. Thus, it is therefore a scientific inquiry and religious intentionality.

3.4 Ethical and Epistemological Discussion

The ethical component of the framework is not a code, but a structural component of the framework. This adoption is opposed to the vocation of AI as neutral in that it operates under the assumption that systems that interact with sacred knowledge are required to exhibit epistemic humility; they ought to be painfully aware of the constraints of computational approaches when working with Quranic writings. The addition of the Ethical and Religious Layer, therefore, serves as a kind of restrictive wall, thus ensuring that all learning activities are based on the necessary characteristics of respect and reverence.

This method is consistent with the modern trends in Responsible AI and moves the discussion beyond the legal and social aspects. It brings about spiritual responsibility as an initial aspect of machine intelligence. The framework adds spiritual epistemology to the field of AI and redefines responsibility as technical and transcendent.

Moreover, this new paradigm adds to the new concept of Faith-Conscious AI. This orientation is actively forming, already expecting a fundamental exploration into how spiritual epistemologies could shape the whole machine intelligence formation and activity. This perspective makes interdisciplinary investigation of interaction of faith-based systems with computational cognitive processes possible.

The Quranic paradigm is introduced in context of the analysis as a paradigmatic example of the concurrent unifying activities of automation and reverence. In the contemporary context of computational philosophy, this synthesis is notable as a rare meeting of technology and spiritual appreciation that can teach useful lessons on how to design and operate intelligent systems.

3.5 Implications for Future Research

Theoretical progress has a variety of future developments, which are:

1. Formalization of Quranic Ontology:

   Theoretical framework of ontological systems that define the principles of Tajweed, Qiraat variations and phonetic articulations. It provides a strong foundation of learning systems on basis of vast knowledge bases.

2. Interdisciplinary Methodological Innovation:

   It is suggested that a collaboration of computer scientists, specialists in the field of Quranic linguistics and theologians would offer a chance to establish a new interdisciplinary field, which can be referred to as Computational Quranology or Digital Quran Computing.

3. Explainable Quranic AI development:

These tools that assist in comprehending the model by visualizing the interaction between the model and the Quranic recitation will be necessary to ensure model transparency and trust in the model by the users.

4. Ethical Norms and Standardization:

   Institutionalization and enforcement of moral principles on AI systems that access sacred texts can assist in the development of model ethical frameworks to inform AI world governance guidelines.

3.6 Limitations and Theoretical Challenges

Concept of proposed paradigm appears to be quite excellent, but its practical implementation should deal with a great number of philosophical and methodological questions.

- **Qiraat Diversity:** The system must operate with the variety of different types of Quranic recitation.
- **Translational ambiguity:** The ability to express theoretical constructs in computational representations and retain all meaning of the construct is an important task.
- **Interdisciplinary Collaboration:** The task is to include the areas of linguistics, technology and theology within one modeling framework.

These challenges should not be viewed as limitations but as possible sources of knowledge, which will push AI research to more profound paradigms.

4. CHALLENGES AND CONSIDERATIONS

The creation of computational tools of the Quranic sciences is a lot of discrete tasks, which presuppose the careful analysis of linguistic, technical and cultural aspects. One of these difficulties is handling various Arabic dialects. The modern Arabic dialects are quite different from one another. However, the Quran recitation is also based on the same set of rules of fixed pronunciation, including the ones of Tajweed, a detailed account of Quranic recitations. The Tajweed rules mostly have an impact on the accurate recitation of the Quran regardless of the accent and dialect of the speaker. It is, consequently, an issue with the computer systems since they must be able to meet the phonetic and prosodic standards of Quranic Arabic in their entirety. They are identical to those of any other accent or dialect of a speaker. The rules are associated with the lengthening of the vowels, nasalization, and emphasis [27]. Developers should implement Tajweed rules into computational systems in order to generate accurate recitation and provide pedagogically meaningful feedback. Without this kind of integration, systems have a possibility of providing phonetically close results that cannot be theologically and pedagogically sound.

Moreover, the integration of Tajweed in the automation processes necessitates the handling of phonetic and acoustic events such as pharyngealization and velarization. They are essential for the right articulation and allow fine gradation between closely related sounds to be distinguishable [4]. The solution to these phonetic difficulties lies in the sophisticated speech recognition methods capable of capturing the delicate acoustic characteristics and being highly precise. Past research has investigated use of enhanced forced alignment techniques to increase phoneme segmentation in Quranic recitation [16].

Apart from the linguistic elements, the availability of computers is still one of the most important aspects of developing these systems. The learning of Quranic recitation demands non-stop real-time feedback. Hence, the systems should be very lightweight, fast, and run in under-resourced areas [28]. The balance between precision, guaranteeing proper pronunciation and Tajweed rules, as well as speed is critical, particularly for mobile and web-apps which are directed at the teaching of the Quran [29, 30].

One of design constraints that continues to be of great concern in Quranic recitation technologies is religious sensitivity. Since the Quran is considered to be the word of Allah, it is deemed to be extremely inappropriate to mispronounce or misrepresent it [31]. To this end, any automated feedback should be informative and non-objectionable. It should be presented in the context of the principles of textual integrity, limited context of use (education/research) and transparency. This encourages incorporation of Tajweed conscious check systems that may detect mispronunciations and give instructions in a respectful, scholar-auditable way [14, 32].

Such technologies have the potential to increase access to Quranic education, especially in under-resourced environments where effective teaching is scarce. This would allow timely and supportive feedback on recitation and understanding to learners in different locations [5]. Nevertheless, to reach this objective, recognition accuracy, cultural sensitivity and educational effectiveness must be constantly improved so that system allows respectful learner-text relationship that is in line with Quranic sciences [13].

Currently, this problem is addressed in a number of directions; some authors focus on enhancing training dataset by offering repositories to developers [33], question-answer datasets [34] and model selection and tuning methods [17]. One of these programs was the development of a CMU Sphinx language model tailored to the Holy Quran, which employed simple Arabic phonemes to enhance speech recognition accuracy in Quranic recitation [35]. El Amrani introduced this method and emphasized on solving the problems regarding Quranic phonetics through the application of a reduced phoneme inventory that facilitates the recognition process to make it more accurate and efficient in dealing with the distinctive phonetic traits of Quranic Arabic [35]. Correspondingly, the study [33] tested the use of large language models for embedding based searches in Quranic texts. The study made a considerable impact on the efficiency of the retrieval of texts and their understanding in context. Parallel to this, the deployment of the latest language models' powers in designing the aforementioned tools aims to support better access to Quranic knowledge through advanced search functionalities. Such developments make the Quran more accessible for educational and scholarly purposes.

In addition, the research [34] on the retrieval of Quran passages optimization confirmed the increasing reliance on proper language models for the enhancement of query-answering systems' accuracy. By developing comprehensive question-answer dataset and language models' fine-tuning, [34] led to the decisive and context-aware retrieval of Quranic passages. This advancement allows scholars and teachers to engage with the text more effectively. This study was also supported by [17], who pointed out the difficulties concerning the use of transformer-based models with Classical Arabic datasets, particularly those based on the Quran and Hadith. On the one hand, transformers seem to be the most efficient models for almost every NLP application. However, their shortcomings in grasping the complexities of Quranic recitation and mastering the syntax of Classical Arabic can tailor strategies for better the performance among the researchers [17]. Moreover, creation of Arabic question-answering corpora like HAQA and QUQA of the Quran and Hadith shows the necessity for large and well-organized datasets that facilitate the development of theoretical and practical computational tools for Quranic studies [36]. Alongside advances in text classification [37] these endeavors are part of the development of automated Quranic recitation and text analysis in computational linguistics.

The development of the language models with aim to detect the Quranic recitation is still in a distant practically achievable form, especially in terms of the proper use of the Tajweed rules. There are still many stumbling blocks that are to be meticulously studied and rectified. The rudimentary models that are generally employed in the general Arabic speech recognition systems are abjectly lacking in respect of their applicability in the conditions of their application in the Quranic recitation.

The issue of recording and standardization of the pronounced phonetic peculiarities of Quranic chanting and performing them with the necessary accuracy is one of the most difficult ones. Tajweed needs to be pronounced precisely. but existing difference is observed in the recitation by various geographical areas and recitations of various persons. Even with the advances in speech recognition and deep learning, the models remain susceptible to retaining correct results in the case of difficult phenomena (vowel lengthening, nasalization, etc.) which are typical of Quranic recitation [29, 32].

The absence of big, good and diversified Quranic speech corpora actually makes the systems that are available unable to achieve accuracy and Tajweed conformity. By this, the design of the universal language models of the Quranic recitations which can fulfill such peculiarities is actually determined. It is issue of implementing the general techniques of speech recognition but in which case there are developed phonetics, linguistic and acoustic characteristics that are peculiar to the Quran reciting. This is the first gap that the researchers and developers will reduce to a minimum due to attention they will pay to the data collection process, enhancement of Tajweed-conscious models and creation of advanced recognition techniques. This is also surpassed by the fact that the accuracy, cultural respect and the religious connotation of the automated systems leads to teaching and preserving recitation of the Holy Quran.

Even though there is an improvement in the situation. the researchers are still unable to create the correct Quranic ASR systems since they have to deal with the phonetic and prosodic complexity needed to comply with Tajweed, and with the lack of large, balanced and well-marked corpora. The limitations cause the need of Tajweed-aware learning and inference algorithms, reproducible benchmarking and conventional dataset design. More research should therefore be done in the development of scalable corpora with standardized Tajweed annotations and assessment procedures. This will enable concurrent measure of phonetic fidelity as well as transcription accuracy and also high educational feedback and cross-model comparisons.

5. CONCLUSION

In order to establish the Quranic speech, this paper suggests deep learning-based model integrating standards of assessment, learning and inference, linguistic-phonetic truth, data structure and religious-ethical control. The framework views Tajweed as internal phonological constraint which defines phoneme-level representation and verification as an alternative to external add-on. To make the technical clarity more understandable, paper provides workflow and quantifiable products (orthographic, phonetic and Tajweed products). These artifacts help in benchmarking in future and repeatable system design. The framework may serve as the bridge between the contemporary AI and the conventional Quranic education. It can preserve reverence, transparency and text integrity and encourage faithful recitation assessment in educational and accessibility scenarios. The primary objectives of future studies should be large-scale annotated corpora, standardized evaluation protocols and explainable Tajweed-conscious models that can be used in practice.

References

[1] Al-Fadhli S, Al-Harbi H, Cherif A. Speech Recognition Models for Holy Quran Recitation Based on Modern Approaches and Tajweed Rules: A Comprehensive Overview. Int J Adv Comput Sci. Appl. 2023;14.

[2] Alagrami AM, Eljazzar MM. Smartajweed: Automatic Recognition of Arabic Quranic Recitation Rules. 2020. ArXiv preprint: https://arxiv.org/pdf/2101.04200

[3] Almousa FA, Al-Mohanna FM. The Quranic Conditionally Pharyngealized Sounds: An Optimality Theory Perspective. AWEJ for Translation & Literary Studies. 2021 Aug;5.

[4] Alsurf, S.S.S. The Phonetics of the Qur'ānic Pharyngealised Sounds: Acoustic and Articulatory Studies. [Phd Thesis]. Macquarie University. 2022. URL: https://figshare.mq.edu.au/articles/thesis/The_Phonetics_of_the_Qur_a_nic_pharyngealised_sounds_acoustic_and_articulatory_studies/19434662?file=34530932

[5] Abro B, Naqvi AB, Hussain A. Coran Recognition for the Purpose of Memorisation Using Speech Recognition Technique. In: Proceedings of the 2012 15th international Multitopic conference (INMIC). IEEE. 2012:30-4.

[6] Mohammed A, Sunar MS, Salam Hj, Md.S. Quranic Verses Verification Using Speech Recognition Techniques. J Teknol. 2015;73.

[7] O'Shaughnessy D. Invited Paper: Automatic Speech Recognition: History, Methods and Challenges. Pattern Recognit. 2008;41:2965-29679.

[8] Balula NO, Rashwan M, Abdou S. Automatic Speech Recognition (Asr) Systems for Learning Arabic Language and Al-Quran Recitation: A Review. IJCSMC. 2021;10(7):91-100.

[9] Abdou SM, Rashwan M. A Computer Aided Pronunciation Learning System for Teaching the Holy Quran Recitation Rules. In: Proceedings of the 2014 IEEE/ACS 11th international conference on computer systems and applications (AICCSA). IEEE.2014:543-550.

[10] Deng L. Deep Learning: From Speech Recognition to Language and Multimodal Processing. APSIPA Trans Signal Inf Process. 2016;5.

[11] Nassif AB, Shahin I, Attili I, Azzeh M, Shaalan K. Speech Recognition Using Deep Neural Networks: A Systematic Review. IEEE Access. 2019;7:19143-19165.

[12] Alrumiah SS, Al-Shargabi AA, Deep A. A Deep Diacritics-Based Recognition Model for Arabic Speech: Quranic Verses as Case Study. IEEE Access. 2023;11:81348-81360.

[13] Shakeel MA, Khattak HA, Khurshid N. Deep Acoustic Modelling for Quranic Recitation – Current Solutions and Future Directions. IPSI Trans Internet Res. 2024;20:61-73.

[14] Harere AA, Jallad KA. Mispronunciation Detection of Basic Quranic Recitation Rules Using Deep Learning. 2023. ArXiv preprint: https://arxiv.org/pdf/2305.06429

[15] Dukes K, Atwell E, Sharaf AB. Syntactic Annotation Guidelines for the Quranic Arabic Dependency Treebank. In: Proceedings of the LREC (language resources and evaluation conference). 2010.

[16] Alqadasi AM, Zeki AM, Sunar MS, Salam MS, Abdulghafor R, et al. Improving Automatic Forced Alignment for Phoneme Segmentation in Quranic Recitation. IEEE Access. 2024;12:229-244.

[17] Altammami S, Atwell E. Challenging the transformer-based models with a classical Arabic dataset: Quran and Hadith. In: Proceedings of the thirteenth language resources and evaluation Conference (LREC 2022). European Language Resources Association (ELRA). 2022:1462-1471.

[18] Mamatov NS, Niyozmatova NA, Abdullaev ShSh, Samijonov AN, Erejepov KK. Speech Recognition Based on Transformer Neural Networks. In: Proceedings of the 2021 international conference on information science and communications technologies (ICISCT). IEEE. 2021:1-5.

[19] Miao H, Cheng G, Gao C, Zhang P, Yan Y. Transformer-Based Online CTC/Attention End-To-End Speech Recognition Architecture. In: Proceedings of the ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE. 2020:6084-6088.

[20] Hadwan M, Alsayadi A, H, AL-Hagree S. An End-To-End Transformer-Based Automatic Speech Recognition for Coran Reciters. Comput Mater Continua. 2023;74:3471-87.

[21] Omran D, Kandil A, ElBialy A, Samy S, fawzy S. CNN for Speech Recognition Case Study: Recitation Rules of the Holy Quran. MSA Eng J. 2023;2:1-12.

[22] Harere AA, Jallad KA.Quran Recitation Recognition Using End-To-End Deep Learning. 2023. ArXiv preprint: https://arxiv.org/pdf/2305.07034

[23] Samara G, Al-Daoud E, Swerki N, Alzu'bi D. The Recognition of Holy Qur'an Reciters Using the MFCCS' Technique and Deep Learning. Adv Multimedia. 2023;2023:1-14.

[24] Saber HA, Younes A, Osman M, Elkabani I. Quran Reciter Identification Using NASNetLarge. Neural Comput Appl. 2024;36:6559-6573.

[25] Issa S, Ayyoub M, Khaleel O, Elmitwally N. Towards Building a Speech Recognition System for Quranic Recitations: A Pilot Study Involving Female Reciters. Jordan J Electr Eng. 2022;8:307.

[26] Al-Issa S, Al-Ayyoub M, Al-Khaleel O, Elmitwally N. Building a Neural Speech Recognizer for Quranic Recitations. Int J Speech Technol. 2023;26:1131-1151.

[27] Mohamed Y, Hoque M, Ismail THSB, Ibrahim MH, Saad NM, Zaidi NN. Relationship Between Phonology, Phonetics, and Tajweed: A Literature Review. In: Proceedings of the 4th International Conference on Sustainable Innovation. Paris, France: Atlantis Press; 2020-Social, Humanity, and Education (ICoSIHESS 2020). 2021.

[28] Abushariah MA, Ainon RN, Zainuddin R, Elshafei M, Khalifa OO. Natural Speaker-Independent Arabic Speech Recognition System Based on Hidden Markov Models Using Sphinx Tools. In: Proceedings of the international conference on computer and communication engineering (ICCCE'10). IEEE. 2010:1-6.

[29] MESSAOUDI F, MANSOURI M, RABAHI H. The Phonetic Miracle in the Quran. ATRAS J. 2025;6:19-33.

[30] Dhouib A, Othman A, El Ghoul O, Khribi MK, Al Sinani A. Arabic Automatic Speech Recognition: A Systematic Literature Review. Appl Sci. 2022;12:8898.

[31] Holy SF. Quranic Manuscripts: Examining Historical Variants and Transmission Methods. Al-Afkar. J Islamic Stud. 2024;7:1163-1177.

[32] Gerhana YA, Atmadja AR, Maylawati DS, Rahman A, Nufus K, et al. Computer Speech Recognition to Text for Recite Holy Quran. IOP Conf Ser Mater Sci Eng. 2018;434:012044.

[33] Alqarni M. Embedding Search for Quranic Texts Based on Large Language Models. Int Arab J Inf Technol. 2024;21.

[34] Basem M, Oshallah I, Hikal B, Hamdi A, Mohamed A. Optimized Quran Passage Retrieval Using an Expanded QA Dataset and Fine-Tuned Language Models. 2024. ArXiv preprint: https://arxiv.org/pdf/2412.11431

[35] El Amrani MY, Rahman MM, Wahiddin MR, Shah A. Building CMU Sphinx Language Model for the Holy Quran Using Simplified Arabic Phonemes. Egypt Inform J. 2016;17(3):305-14.

[36] Alnefaie S, Atwell E, Alsalka MA. HAQA and QUQA: Constructing Two Arabic Question-Answering Corpora for the Quran and Hadith. In: Proceedings of the proceedings of the conference on recent advances in natural language processing (RANLP 2023). 2023:90-97.

[37] Rostam NA, Malim NH. Text Categorisation in Quran and Hadith: Overcoming the Interrelation Challenges Using Machine Learning and Term Weighting. J King Saud Univ Comput Inf Sci. 2021;33:658-67.