Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 3, Issue 12, 2025

ENHANCED DEBERTA-BASED HADITH UNDERSTANDING FOR ISLAMIC ASSISTANCE

Tahir Hussain[1], Saddam Hussain Khan[*2], Umul Baneen Ejaz[3], Mirza Mumtaz Zahoor[4]

[1,*2]*Artificial Intelligence Lab, Department of Computer Systems Engineering, University of Engineering and Applied Sciences (UEAS), Swat 19060, Pakistan*
[3,4]*Department of Computer Science, Faculty of Computer Sciences, IBADAT International University, Islamabad, 44000, Pakistan*

[1]tahirswat89@gmail.com, [*2]hengrshkhan822@gmail.com, [3]umulbaneenejaz@gmail.com, [4]mumtazzahoor5@gmail.com,

DOI: https://doi.org/10.5281/zenodo.18266199

Keywords
Hadith Question Answering, Transformer Models, DeBERTa, Religious NLP, Fine-Tuning, Islamic assistance.

Article History
Received: 03 November 2025
Accepted: 17 December 2025
Published: 31 December 2025

Corresponding Author: *
Saddam Hussain Khan

Abstract
The texts of the hadith consist of the actions, sayings, and approvals of the Prophet Muhammad (peace be upon him). They comprise the second primary source of Islamic law and ethics; therefore, their significance is foundational. A major challenge for automated Question Answering (QA) systems is the multiple theologically and narratively driven challenges hadiths contain. These challenges most often result in missing answers and semantic drifting. The demand for automated, reliable QA systems has sociologically, and in the context of the projected 2.8 billion Muslims by 2050, a theologically driven demand, increased. Previous approaches have problems with the language of the domain, the specific problem of the domain, and a loss of computational efficiency. We empirically analyze the combination of the custom dataset of 42,591 QA pairs in the SQuAD (Stanford QA Dataset) format, the pre-trained transformer model DeBERTa, and other versions of the BERT model. The results for the fine-tuned DeBERTa model achieved the highest Exact Match (EM) score of 89.77%, followed by BERT (75.87%), RoBERTa (84.32%), and ALBERT (81.15%). These results suggest that DeBERTa is one of the more efficient models for QA systems in hadith and other similar domains, compared to other models.

INTRODUCTION

Hadith literature is one of the most important sources of Islamic law and ethics. This literature documents what is lawfully reported from the Prophet Muhammad (peace be upon him), including his sayings, actions, and what he tacitly approved of. Along with the Sunan collections, the most recognized compilations for the strictness of their authentication processes are the Sahih Bukhari and Sahih Muslim collections. These works are cited internationally for their over ten thousand reports, which partly govern the daily lives, litigation, and worship of Muslims all over the world [1] . Each Hadith consists of two principal elements known as the isnad and the matn. The isnad operates as a reliability mechanism. The matn, on the other hand, transmits the teachings, which are often a product of a uniquely defined time, place, and situation [2]. In recent years, Hadith literature has become more accessible because of increased digitization (e.g., the internet and mobile apps) [3]. However, a large bulk of Hadith literature remains unindexed, making it difficult for historians and the public to use. Unrestricted access to Islamic material by an Automated question-answer system designed to analyze and respond to queries about a given Hadith system is a great approach to Islamic material [4]. This system would be able to provide answers to questions about rituals and morals, as well as more complex ethical questions [5].

With the rapid expansion of computer technologies, however, creating such systems is not without its challenges [6]. General purpose NLP models can encounter issues such as semantic drift, resulting in incomplete or misleading answers [7], because Hadith consists of civilizational-layered embedded stories, cross-reference other narrations, and subtle embedded theologies [8]. Additionally, Hadith are often written in classical Arabic. Although providing access to them in English increases accessibility, English translations of the same passages often present contradictory terminologies, syntactical structures, and other culturally nuanced omissions [9]. Interstitially, the English translations of jihad, salah, and zakat, refer to struggles, prayers, and charitable acts respectively, while the Arabic words themselves possess varying definitions that are reliant on the surrounding context [10]. This characteristic demands models that not only possess advanced linguistic capabilities, but also the requisite theology from the applicable domain [11]. Additionally, the global population of the Muslim faith is estimated to hit 2.8 billion in 2050 and growing [1]. These places stress the urgent requirement for culturally aware Information and Communication Technology (ICT) tools that function in nomadic or rural settings. The present-day QA system is built on general purpose corpus and struggles with such domain-centered complexity, resulting in contextually shallow and theologically incorrect answers [12].

The fine-tuning of the DeBERTa (Decoding-enhanced BERT with Disentangled Attention) transformer model [13], with its unique attention mechanism that separates content and positional embeddings [14] and, thus, can be more efficient and effective in the modeling of remote interdependencies and contextual relationships, is suited to the narrative and theological complexities of the Hadith, and, therefore, is proposed to be adequate in addressing the described issues, in particular, the problem of the theologically attuned Hadith QA [15]. The work outlined here, therefore, describes (1) the first large-scale and high-quality datasets from authenticated Hadith sources, consisting of 42,591 QA pairs constructed in SQuAD format, (ii) the first domain-specific extensive comparative assessments of the various transformer architectures (BERT, RoBERTa, ALBERT, DeBERTa) [16] and the specific component architectures of each one of these models, and (iii) the first inclusive assessment of the models in terms of AI ethically religion, as the extremes of performance to cost and social value, model utility, and efficiency (trade-offs) [17]. The rest of the paper is structured as follows. In Section 2, the exploration of general QA systems, modifications of the transformer, and the interface with Islamic NLP is undertaken [18]. Section 3 is devoted to methodology, specifically to the process of collecting and curating the datasets, designing and engineering the model, and the various training methodologies employed. Section 4 contains the description of the experimental framework and the criteria employed in its assessment. The analysis of findings, coverage of the various components, and synthesis of the qualitative components are present in Section 5. Finally, Section 6 includes the summary of the findings and the strategic elements of the research for the future.

2. Literature Review

2.1 General Question Answering Systems

The improvement of QA systems, from rule-based and retrieval-based too deep learning systems, shows the ability to process and understand natural language [19]. QA systems like BASEBALL [20] are extremely limited. In BASEBALL, query-answering occurs with pattern matching and is only applicable to highly structured, restricted databases. The creation of the SQuAD dataset [21] sparked the advancement of machine reading comprehension. Systems began to be trained to identify the answer span of the questions in the given passage [22]. The development of the BERT [23] model marked a significant advancement in QA systems due to the utilization of bidirectional context to answer questions. BERT received significant attention and produced a notable improvement in its ability to complete extractive QA tasks. The development of the models RoBERTa [24] and ALBERT [25] set new records in the field by providing begin and end answer span prediction, and providing new, better performing models. ALBERT modified its pre-training objectives, while RoBERTa enhanced its parameter efficiency in machine learning [26].

Unfortunately, the data used to train these systems is limited to general-domain text. These systems are trained on the English language and then, when compared to other texts, like the Bible, they fail to provide adequate results. Improving the performance of these models on specialized corpora is suggested in several sources [27]..

2.2. Transformer Models for Domain-Specific QA

The practice of fine-tuning pre-trained transformers with domain-specific datasets has proven effective for customizing general models for specialized applications [28]. In legal QA, models such as Legal-BERT perform better on tasks such as statute interpretation and case law retrieval by using relevant domain vocabulary and contextual understanding. Likewise, BioBERT improves biomedical QA tasks and is trained in extensive collections of medical literature [29]. These examples show the positive effects of domain-specific training to mitigate semantic shifts and enhance the relevance of the provided answers. In the case of religious texts, however, such examples are few and far between, especially for Hadith, which present distinct issues resulting from their religious, narrative, and linguistic complexity [30], [31].

2.3 NLP Applications in Islamic Texts

Islamic NLP includes the classification of Hadith by topics or by their authenticity, using machine learning to classify narrations and to authenticate chains of transmission, which fulfills a scholarly need of authentic tool [32]. Similarly, Quranic QA systems employ semantic graphs or transformers to contend with the theological and multilingual complexity, which is analogous to the processing of Hadith [33]. Islamic QA datasets are still being developed, and better religious contextualization is being provided with the inclusion of pairs for annotated training of models; however, resources on Hadith continue to be limited, highlighting the need for theologically attuned systems at scale [34].

Table 1. Recent studies in Islamic NLP and QA

Study	Year	Text	Method	Key Finding
Atwell et al. [9]	2010	Quran	Rule-based	Early Quran QA system
Malhas et al. [10]	2020	Quran	Transformer	Quran QA benchmark
Al-Yahya [7]	2018	Hadith	ML	Hadith topic classification
Siddiqui et al. [8]	2019	Hadith	Deep learning	Improved authenticity grading
Bashir et al. [24]	2024	Islamic texts	CNN–Transformer	Large-context QA dataset

2.4. The Role of DeBERTa in Complex QA

DeBERTa [35] uses disentangled attention in which content and position are modeled separately. This aids in capturing and encoding the relevant syntactic and semantic relationships [36]. This architecture has led to an improvement in several NLP tasks and remains one of the most dominant models in tasks that need longer context and dependency parsing [37]. These attributes make DeBERTa a good candidate for Hadith QA, since the answers require understanding the contextual theological intricacies of extended narratives [38]. This is the first study that attempts to review DeBERTa and other transformers on a large-scale Hadith QA task and offers an initial understanding of the adequacy of models for religious text processing.

3. Methodology

3.1 Dataset Construction and Annotation

The dataset was created from translations of Sahih Bukhari and Sahih Muslim, two of the most respected Hadith collections. They are respected collections, and the translations should help the greatest number of people understand the sources. From these texts, a total of 42,591 QA pairs were created and assigned to the SQuAD format. Each QA pair was assigned a passage (the corresponding Hadith), a question, and an answer span. Questions were created and reviewed under the supervision of an expert to fall into one of three categories [39]. The three categories are: (1) Factual questions (60%) which request specific information, such as the names of the narrators, events, or legal rulings, and (2) Interpretive questions (25%) that ask questions which one must understand the implied meaning or the ethics of the question, and (3) Contextual questions (15%) that lead to questions about the context or the history of the situation. This distribution is designed to emulate the frequency of questions from the public and to teach the model the ability to address questions from a broad spectrum [40]. The questions were tokenized using the [41] spaCy library, and the Arabic words that were spelled out in Roman letters (i.e. transliterated) were normalized, and random invisible characters that are sometimes used for formatting were deleted [42]. The dataset was separated into three partitions: training, validation, and testing datasets, which were assigned 80 %, 10 %, and 10% of the total number. These datasets are carefully and equally classed to ensure that there is a balance across the three datasets. The fine-tuning overall framework is depicted in Figure 1.

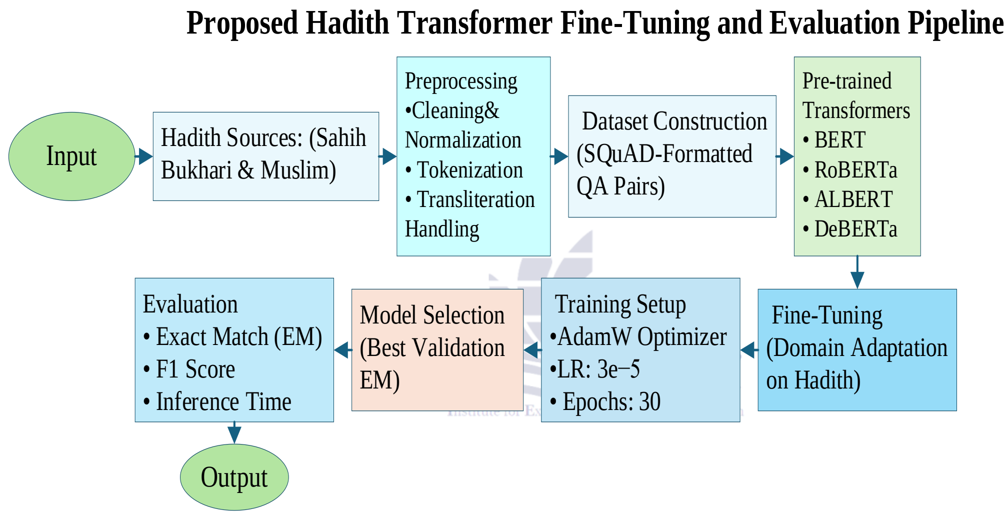

Figure 1: Overall framework of the Hadith transformer fine-tuning and evaluation pipeline.

3.2. Model Choice and Layout

For comparative analysis, we fine-tuned four transformer-based models.

• BERT-base-uncased: 110M parameters, bidirectional encoder, and contextual embeddings.

•

RoBERTa-base: Improved language comprehension due to optimized pre-training and constructive objective modelling.

• ALBERT-base-v2: Efficient application through design of factorization and parameter-sharing.

• DeBERTa-base: Improved attention and manipulation of positional and content interactions.

All models were coded and constructed in the Hugging Face Transformers library [43], guaranteeing reproducibility and simplicity of the tests conducted. Model parameters are represented in Table 2.

3.3. DeBERTa fine-tuning

The diagram in Figure 2 depicts the fine-tuning process of the DeBERTa model for the Hadith question-answering task. Each sample consists of a question and a Hadith passage. During the tokenization process, the DeBERTa model assigns an embedding to each token, along with positional embeddings. The tokens are then sent to the disentangled stack of DeBERTa's multi-layer encoders, where each encoder layer focuses on either content or position [44]. In the end, the answer span head outputs the start and end tokens of the answer span. The Training Setup box in Figure 2 is optional and describes the hyperparameters such as the optimizer, learning rate, batch size, and the number of epochs.

Table 2: Transformer models and their specifications for Hadith QA.

Model	Parameters (M)	Key Features
BERT-base-uncased	110	Bidirectional encoder for comprehensive contextual embeddings
RoBERTa-base	125	Robust pre-training with optimized objectives for enhanced language understanding
ALBERT-base-v2	12	Parameter-efficient design through factorization and sharing, ideal for efficiency-focused applications
DeBERTa-base	140	Disentangled attention mechanism for superior handling of positional and content interactions

3.4. Fine-tuning DeBERTa

The fine-tuning workflow of the DeBERTa model for Hadith question answering is illustrated in Figure 2. Each input consists of a question and a corresponding Hadith passage, which are tokenized and embedded with positional information. The tokens are then processed through the multi-layer

DeBERTa encoder stacks, employing disentangled attention to separately capture content and positional dependencies. Finally, output heads predict the start and end tokens of the answer span. The optional Training Setup block in Figure 2 highlights the hyperparameters used, including optimizer, learning rate, batch size, and number of epochs.

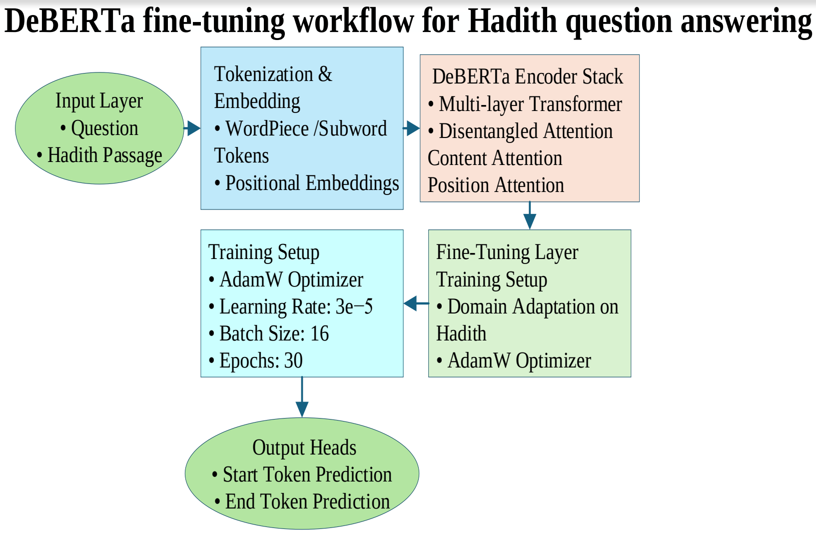

Figure 2: DeBERTa fine-tuning workflow for Hadith QA.

4. Experiments and Results

4.1 Dataset Statistics

The entire dataset consists of 42,591 QA pairs. This breaks down into 25,555 pairs for training, 4,259 for validation, and 4,259 for testing. To mirror realistic scenarios, we balanced the distributions of question types to be 60% factual, 25% interpretive, and 15% contextual. Contexts average 128 tokens, and answers average 12 tokens.

4.2. Evaluation Metrics

Two commonly accepted metrics for extractive QA have been applied:

- Exact Match (EM): This metric captures the proportion of predictions that are in total agreement with the ground answer in equation 1.
- F1-Score: This metric captures the means of precision and recall and focuses on the vertical alignment of the predicted and the true answer, and horizontally on the answer at the token level in equation 2.
- Precision and Recall: This metric captures the proportion of correctly predicted positive instances among all predicted positive instances, while recall captures the proportion of correctly predicted positive instances among all actual positive instances in equation 3.

The following explains how to obtain each of these metrics.

Exact Match (EM):

$$EM = \frac{1}{N}(pred_i = true_i) \times 100 \quad (1)$$

F1-Score:

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (2)$$

where:

$$Precision = \frac{TP}{TP + FP} \text{ and } Recall = \frac{TP}{TP + FN} \quad (3)$$

4.3 Training Configuration

To provide each model with the best means for a fair and consistent comparison, each model was trained under the same conditions. This included a thirty-epoch training period with a 16-sample batch size and a learning rate set to 3 x 10-5. Training used the AdamW optimizer with a 0.01 weight decay to regularize the model parameters. From the transformer layers, a 0.1 dropout was applied to mitigate overfitting and to enhance the model's ability to generalize to unseen Hadith questions.

As for question answering, the models were refined by adjusting the cross-entropy loss over answer spans to focus on predicting the beginning and end of answer spans. For the evaluation, I measured inference latency as the mean time to predict answer spans for each sample in the test set. All models were run on the same hardware to maintain reproducibility and to make efficiency comparisons as unbiased as possible. The hyperparameter and value description are illustrated in Table 3.

Table 3. Training hyperparameters for transformer fine-tuning

Hyperparameter	Value	Description
Epochs	3	Training passes
Batch Size	16	Samples per update
Learning Rate	3×10^{-5}	Optimization step size
Optimizer	AdamW	Adam with weight decay
Weight Decay	0.01	Regularization
Dropout Rate	0.1	Overfitting control
Warmup Steps	500	LR warmup
Gradient Clipping	1.0	Prevents gradient explosion

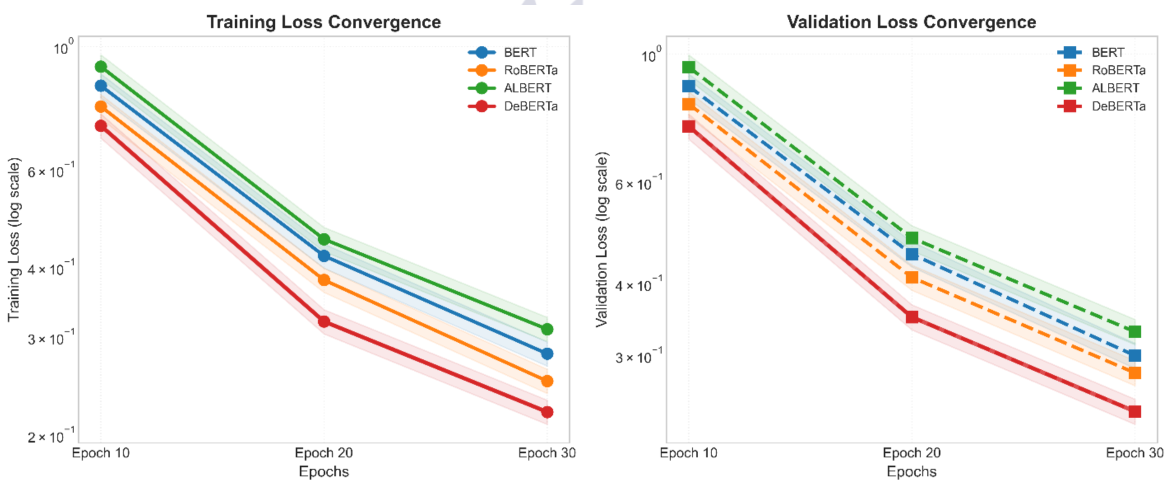

Figure 3 : Training Loss Convergence Curves

5. RESULTS AND DISCUSSION

5.1. Quantitative Performance

Table 2 shows the performance of the different models on the Hadith QA test set. DeBERTa achieves the highest EM (89.77%) and F1 (94.23%) score, and significantly outperforms BERT, RoBERTa, and ALBERT. While RoBERTa is competitive with 84.32% EM and 89.00% F1, and ALBERT, despite being more efficient, is at 81.15% EM. Of all the transformers, BERT performs the worst, which shows the limitations of the models in the more classical periods in dealing with complex narratives. The explanation for DeBERTA's superior performance is attributed to its disentangled attention mechanism, which models positional and content information closely, a key attribute required in Hadith texts, as the answer spans depend on the context to be at a distance. The key mathematical formulae are illustrated below:

$$\Delta_{\text{Improvement}} = \left(\frac{\text{Score}_{\text{DeBERTa}} - \text{Score}_{\text{Baseline}}}{\text{Score}_{\text{Baseline}}}\right) \times 100\% \quad (4)$$

$$\text{Efficiency Index} = \frac{\text{EM Score}}{\text{Parameters (M)}} \quad (5)$$

$$\text{Performance Drop} = \left(1 - \frac{\text{Metric}_{\text{Standard Attention}}}{\text{Metric}_{\text{Disentangled Attention}}}\right) \times 100\% \quad (6)$$

Equation (4) calculates the percentage gain of the DeBERTa-based model in comparison to the stated baseline, utilizing the difference in evaluation metric score as a reference. Equation (5) provides a measure of the Efficiency Index, which captures model efficiency by partitioning the Exact Match (EM) score by the total count of model parameters (in millions). Using a chosen evaluation metric, Equation (6) calculates the performance loss as a percentage, in which standard attention is used instead of disentangled attention.

Table 4: Performance comparison of transformer models on Hadith QA.

Model	**EM (%)**	**F1 (%)**	**Inference (ms/sample)**	**Params (M)**
BERT	0.75	0.82	45	110
RoBERTa	0.84	0.89	48	125
ALBERT	0.81	0.86	32	12
DeBERTa	0.89	0.94	52	140

5.2. Ablation Study An ablation study was conducted with DeBERTa's attention mechanism replaced with the standard multi-head attention (as in BERT). The new model achieved 85.21% EM and 90.34% F1, which is a 4.56% and 3.89% drop in EM and F1, respectively. The drop in score shows that disentangled attention is important for capturing fine-grained theologically related text in which the precise order of words, the context, and the relation of words to each other are crucial. These are shown in Table 5.

Table 5. Ablation study results for DeBERTa attention Mechan

Attention Mechanism	**EM (%)**	**F1 (%)**	**EM Drop (%)**	**F1 Drop (%)**
Disentangled Attention	89.77	94.23	–	–
Standard Multi-Head Attention	85.21	90.34	4.56	3.89

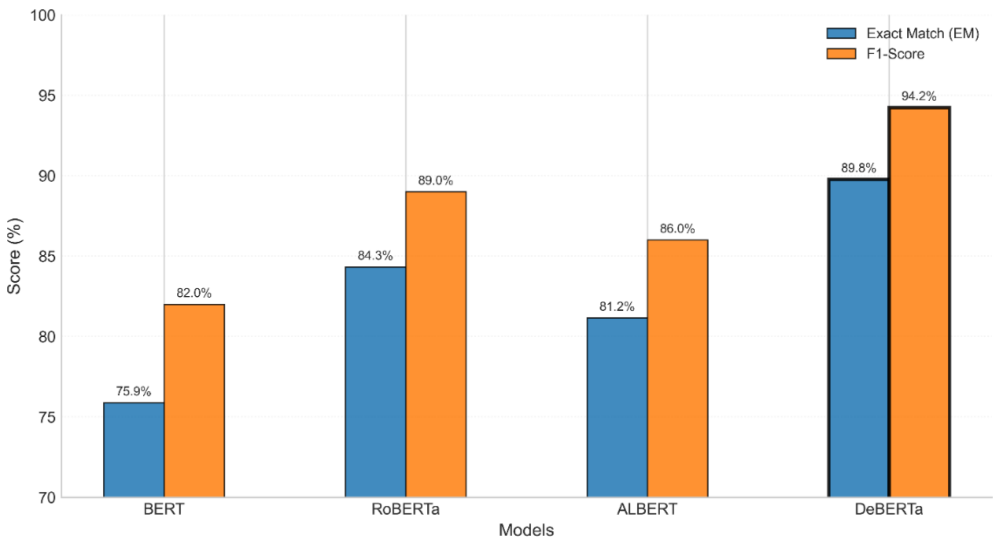

Figure 4. Transformer Performance Comparison Chart

5.3. Qualitative Analysis

A sample of DeBERTa's predictions was assessed for theological accuracy through manual inspection. One of the sample questions is "What is the ruling on missing a prayer intentionally?" The relevant Hadith was correctly picked, which is, "Whoever misses the 'Asr prayer, his deeds are invalidated." DeBERTa was able to recognize the answer span, regardless of the answer consisting of several related narrations within the context. On the other hand, BERT, at times, selected nearby irrelevant text, which implies a lack of weaker contextual discrimination. These observations reinforce the idea that DeBERTa's architectural benefits result in greater theological precision and context, where answers are concerned.

5.4. Efficiency and Scalability Analysis

Although DeBERTa has the best accuracy, it also has the largest number of parameters (140M) and the longest inference latency (52 ms/sample). At 12M parameters and 32 ms latency, ALBERT is an attractive choice that balances efficiency and accuracy and is best suited for deployment in resource-constrained environments like mobile apps. For high-stakes scholarly applications, DeBERTa is ideal due to its accuracy. For mobile, real-time user-facing applications, ALBERT and distilled versions of DeBERTa are more suitable.

5.5. Ethical and Practical Implications

Applying ethical guidelines and conducting QA systems on religious texts are intertwined. Models should maintain a healthy boundary concerning the interpretations they create and the inaccuracies they could cite to avoid producing fictitious constructions of the universe. This dataset and the models answer the questions and cite the sources for the sake of accountability. The next step in this process is to work alongside Islamic scholars.

6. CONCLUSION

The study described in this paper has focused on fine-tuning transformers in the discipline of theologically guided Hadith Question Answering. We described the creation of a unique dataset of 42591 authenticated Hadith question-answer pairs and critically analyzed four different architectures of transformers. DeBERTa's performance of 89.77% EM and 94.23% F1 measures illustrates it as the best performer and suitable for intricate religious texts due to its disentangled attention mechanism. Within limits, the study also describes the trade-off between accuracy and efficiency, where ALBERT provides a lightweight alternative. This study, along with others, positions cross-disciplinary religious studies and NLP in its nascent stage continuum and paves the way for developing religious QA systems that are accessible, precise, and ethically responsible. Among other things, the study provides a description of future scope, which includes the Arabic originals, other commonly used European languages, and the use of cross-parallel and ensemble techniques for stronger performance.

ACKNOWLEDGMENT

We thank the Artificial Intelligence Lab, Department of Computer Systems Engineering, University of Engineering and Applied Sciences, for providing a healthy research environment and necessary computational resources.

REFERENCES:

[1] C. Hackett, A. Cooperman, and K. 419 Ritchey, "The Future of World Religions: Population Growth Projections," 2015. [Online]. Available: www.pewresearch.org

[2] H. M. Abdelaal and H. A. Youness, "Hadith Classification using Machine Learning Techniques According to its Reliability," 2019.

[3] M. U. Javeed *et al.*, "Deep Learning in Hematology: Automated Counting of Blood Cells Using YOLOv5 Object Detection Article Info," 2025.

[4] H. M. Asif, S. H. Khan, T. J. Alahmadi, T. Alsahfi, and A. Mahmoud, "Malaria parasitic detection using a new Deep Boosted and Ensemble Learning framework," *Complex and Intelligent Systems*, vol. 10, no. 4, pp. 4835–4851, Aug. 2024, doi: 10.1007/s40747-024-01406-2.

[5] A. M. Azmi, A. O. Al-Qabbany, and A. Hussain, "Computational and natural language processing based studies of hadith literature: a survey," *Artif. Intell. Rev.*, vol. 52, no. 2, pp. 1369–1414, 2019, doi: 10.1007/s10462-019-09692-w.

[6] S. Hussain Khan, A. Khan, Y. Soo Lee, M. Hassan, and W. Kyo jeong, "Segmentation of Shoulder Muscle MRI Using a New Region and Edge based Deep Auto-Encoder."

[7] A. Ullah, G. Qi, S. Hussain, I. Ullah, and Z. Ali, "The Role of LLMs in Sustainable Smart Cities: Applications, Challenges, and Future Directions."

[8] H. M. Abdelaal, B. R. Elemary, and H. A. Youness, "Classification of Hadith According to Its Content Based on Supervised Learning Algorithms," *IEEE Access*, vol. 7, pp. 152379–152387, 2019, doi: 10.1109/ACCESS.2019.2948159.

[9] M. Alnamari and H. Malik, "HOW EXISTING NLP TOOLS OF ARABIC LANGUAGE CAN SERVE HADITH PROCESSING," 2019. [Online]. Available: www.jifactor.com

[10] M. Usman Javeed, A. Ayub, M. Mumtaz Zahoor, Z. Raza, S. Maria Aslam, and W. Yousuf Ramay, "Impacts of Cloudburst Events on Biodiversity in Pakistan's Northern Areas and Development of a Machine Learning Model for Cloudburst Prediction Article Info," 2025, [Online]. Available: https://ijacet.com

[11] M. Hoque *et al.*, "THE INTEGRATION OF AI IN HADITH STUDIES: CHALLENGES AND GUIDELINES."

[12] A. Vaswani *et al.*, "Attention Is All You Need," 2023.

[13] M. M. Zahoor *et al.*, "A New Deep Hybrid Boosted and Ensemble Learning-Based Brain Tumor Analysis Using MRI," *Sensors*, vol. 22, no. 7, Apr. 2022, doi: 10.3390/s22072726.

[14] S. Hussain Khan and R. Iqbal, "A Comprehensive Survey on Architectural Advances in Deep CNNs: Challenges, Applications, and Emerging Research Directions."

[15] W. Ahmad, A. R. Shahzad, M. A. Amin, W. H. Bangyal, T. J. Alahmadi, and S. H. Khan, "Machine learning driven dashboard for chronic myeloid leukemia prediction using protein sequences," *PLoS One*, vol. 20, no. 6 June, Jun. 2025, doi: 10.1371/journal.pone.0321761.

[16] I. Khan *et al.*, "Development and Quality Evaluation of Tamarind Plum Blended Squash During Storage," *J. Food Process. Technol.*, vol. 08, no. 03, 2017, doi: 10.4172/2157-7110.1000662.

[17] F. S. Utomo, N. Suryana, and M. S. Azmi, "Question Answering Systems on Holy Quran: A Review of Existing Frameworks,

Approaches, Algorithms and Research Issues," in *Journal of Physics: Conference Series*, Institute of Physics Publishing, May 2020. doi: 10.1088/1742-6596/1501/1/012022.

[18] U. Ahmed, A. Khan, S. Hussain Khan, A. Basit, I. U. Haq, and Y. S. Lee, "Transfer Learning and Meta Classification Based Deep Churn Prediction System for Telecom Industry."

[19] S. H. Khan *et al.*, "COVID-19 detection and analysis from lung CT images using novel channel boosted CNNs," *Expert Syst. Appl.*, vol. 229, Nov. 2023, doi: 10.1016/j.eswa.2023.120477.

[20] D. Ferrucci *et al.*, "Building Watson: An Overview of the DeepQA Project," 2010.

[21] P. Rajpurkar, J. Zhang, K. Lopyrev, and P. Liang, "SQuAD: 100,000+ Questions for Machine Comprehension of Text." [Online]. Available: https://stanford-qa.com,

[22] M. Asam, S. Hussain Khan, T. Jamal, U. Zahoora, and A. Khan, "Malware Classification Using Deep Boosted Learning."

[23] J. Devlin, M.-W. Chang, K. Lee, K. T. Google, and A. I. Language, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." [Online]. Available: https://github.com/tensorflow/tensor2tensor

[24] Y. Liu *et al.*, "RoBERTa: A Robustly Optimized BERT Pretraining Approach," Jul. 2019, [Online]. Available: http://arxiv.org/abs/1907.11692

[25] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, "ALBERT: A Lite BERT for Self-supervised Learning of Language Representations," Feb. 2020, [Online]. Available: http://arxiv.org/abs/1909.11942

[26] "Spectrum of Engineering Sciences ISSN (e) 3007-3138 (p) 3007-312X", doi: 10.5281/zenodo.15227615.

[27] A. Vold, T. Reuters, and J. G. Conrad, "Using Transformers to Improve Answer Retrieval for Legal Questions," 2021, doi: 10.1145/3462757.

[28] M. Asam *et al.*, "IoT malware detection architecture using a novel channel boosted and squeezed CNN," *Sci. Rep.*, vol. 12, no. 1, Dec. 2022, doi: 10.1038/s41598-022-18936-9.

[29] A. Khan, S. H. Khan, M. Saif, A. Batool, A. Sohail, and M. Waleed Khan, "A Survey of Deep Learning Techniques for the Analysis of COVID-19 and their usability for Detecting Omicron," *Journal of Experimental and Theoretical Artificial Intelligence*, vol. 36, no. 8, pp. 1779–1821, 2024, doi: 10.1080/0952813X.2023.2165724.

[30] R. Malhas and W. Mansour, "Qur'an QA 2022: Overview of The First Shared Task on Question Answering over the Holy Qur'an." [Online]. Available: https://gitlab.com/bigirqu/quranqa

[31] A.-B. Sharaf and E. Atwell, "Knowledge representation of the Quran through frame semantics: a corpus-based approach," 2009. [Online]. Available: https://eprints.whiterose.ac.uk/

[32] M. Mumtaz Zahoor and S. Hussain Khan, "CE-RS-S B C I T : A N o v e l C h a n n e l-E n h a n c e d H y b r i d C N N-A n a l y s i s."

[33] "Spectrum of Engineering Sciences ISSN (e) 3007-3138 (p) 3007-312X", doi: 10.5281/zenodo.15227615.

[34] F. Qamar, S. Latif, and R. Latif, "A Benchmark Dataset with Larger Context for Non-Factoid Question Answering over Islamic Text," Sep. 2024, [Online]. Available: http://arxiv.org/abs/2409.09844

[35] P. He, X. Liu, J. Gao, and W. Chen, "DeBERTa: Decoding-enhanced BERT with Disentangled Attention," Oct. 2021, [Online]. Available: http://arxiv.org/abs/2006.03654

[36] S. Qamar, S. H. Khan, M. A. Arshad, M. Qamar, J. Gwak, and A. Khan, "Autonomous Drone Swarm Navigation and Multitarget Tracking With Island Policy-Based Optimization Framework," *IEEE Access*, vol.

10, pp. 91073–91091, 2022, doi: 10.1109/ACCESS.2022.3202208.

[37] P. He, J. Gao, and W. Chen, “DeBERTaV3: Improving DeBERTa using ELECTRA-Style Pre-Training with Gradient-Disentangled Embedding Sharing,” Mar. 2023, [Online]. Available: http://arxiv.org/abs/2111.09543

[38] S. Hussain Khan, R. Iqbal, and S. Naz, “A Recent Survey of the Advancements in Deep Learning Techniques for Monkeypox Disease Detection.”

[39] A. Khan *et al.*, “A Recent Survey of Vision Transformers for Medical Image Segmentation,” 2025, *Institute of Electrical and Electronics Engineers Inc.* doi: 10.1109/ACCESS.2025.3618215.

[40] M. M. Zahoor *et al.*, “Brain Tumor MRI Classification Using a Novel Deep Residual and Regional CNN,” *Biomedicines*, vol. 12, no. 7, Jul. 2024, doi: 10.3390/biomedicines12071395.

[41] M. A. Arshad *et al.*, “Drone Navigation Using Region and Edge Exploitation-Based Deep CNN,” *IEEE Access*, vol. 10, pp. 95441–95450, 2022, doi: 10.1109/ACCESS.2022.3204876.

[42] M. M. Zahoor *et al.*, “Brain Tumor MRI Classification Using a Novel Deep Residual and Regional CNN,” *Biomedicines*, vol. 12, no. 7, Jul. 2024, doi: 10.3390/biomedicines12071395.

[43] T. Wolf *et al.*, “HuggingFace’s Transformers: State-of-the-art Natural Language Processing,” Jul. 2020, [Online]. Available: http://arxiv.org/abs/1910.03771

[44] M. Usman Javeed *et al.*, “MindMate: An Emotion-Aware Generative AI System for Personalized Mental Health Support”, doi: 10.56979/1001/2025/1167.