

Technical Note

Impact of Machine Learning Integration in Qur'anic Studies

Arshad Iqbal¹ , Shabbir Hassan^{2, *} 

¹K. A. Nizami Centre for Quranic Studies, Aligarh Muslim University, Aligarh, India

²Department of Computer Science, Aligarh Muslim University, Aligarh, India

Abstract

The advancement in the field of computer science, especially in machine learning (ML), represents a flourishing innovation that carries great importance in the domain of education. The beneficial impact of ML can also be observed in the realm of Qur'anic studies, particularly in Arabic text recognition and recitation analysis. This paper presents a comprehensive analysis of 34+ published scholarly articles devoted to Qur'anic studies. This work explores the convergence of machine learning methodologies and Qur'anic studies, examining the innovative applications and methodologies for Arabic text and voice classification. The fusion of ML algorithms makes the work easy and accurate to analyze, interpret, and extract valuable insights from the sacred text. Subsequently, we delve deeper into the emergent field of ML algorithms like k-NN, ANN, BLSTM, MFCC, SVM, NB and DL approaches have been adapted for Qur'anic texts classification, recitation and recitation analysis on accuracy, speed, class recognition, response rate and biasness benchmark. This work covers a diverse range of applications, including automated Qur'anic exegesis and analysis of usage of Ahkam Al-Tajweed. The main contribution of the work is to provide insight into how ML facilitates in Arabic and Kufic textual analysis, linguistic subtleties, and thematic structures of the Qur'anic text. Using the deep learning approaches, the reciters, recitation style and of the Quranic text has also explained in the work.

Keywords

Hadith Classifier, Kufic Text, Machine Learning, Quranic Recitation, Quranic Text Recognition, Reciter Classification, Tajweed Rules

1. Introduction

Throughout the course of history, humans have consistently endeavored to procure instruments with the objective of enhancing the effectiveness of various tasks in their progress. The limitless creativity of individuals has given rise to the digital epoch, wherein the realm of computer science permeates every aspect of human existence. The impact of computers on daily life is increasing at an exponential rate with each passing year, and it is expected that this trajectory will persist in the foreseeable future. Given the importance of

integrating Artificial Intelligence (AI) in the Islamic World, where it presents intelligent models across diverse applications, this paper undertakes a brief examination of a particular AI methodology—machine learning. The progression of digitalization has sparked numerous alterations, with one of the most notable being the advancement of Machine Learning (ML). ML represents an innovative paradigm that imparts knowledge to machines regarding more efficient methods of processing data [1]. In the rapidly evolving landscape of

*Corresponding author: shassan.cs@amu.ac.in (Shabbir Hassan)

Received: 3 September 2024; **Accepted:** 6 October 2024; **Published:** 29 October 2024



Copyright: © The Author (s), 2024. Published by Science Publishing Group. This is an **Open Access** article, distributed under the terms of the Creative Commons Attribution 4.0 License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

technological advancements, the integration of cutting-edge tools has permeated various fields, enriching traditional domains with novel methodologies. One such intriguing intersection lies at the nexus of Machine Learning (ML) and Qur'anic Studies, where computational methods are employed to delve into the profound wisdom encapsulated in the Holy Qur'an [2]. The Holy Qur'an, considered the foundational text in Islam, and has long been a subject of meticulous study and interpretation. The infusion of machine learning techniques brings forth new possibilities for scholars and researchers to uncover hidden insights, linguistic nuances, and thematic structures within the sacred verses [3]. Recent years have witnessed a surge in research endeavors at the crossroads of machine learning and Qur'anic Studies, marking a paradigm shift in how scholars approach the analysis and interpretation of religious texts [4].

The Qur'an, revealed to Prophet Muhammad, over a period of 23 years, starting when he was 40 years old, serves as the foundational scripture for nearly two billion Muslims globally, constituting approximately 25% of the world's population [5]. After the Prophet's passing, the Qur'an, As-Sunnah (traditions of the Prophet), and ancient Islamic texts have become the primary sources of knowledge, wisdom, and legal guidance for followers of Islam [6]. As a result, the Qur'an holds immense importance in the daily lives of believers. It holds a paramount position as the principal text for Muslims [7], comprising thirty chapters and 6,236 verses organized into 114 groups known as "Surahs." The Qur'an was revealed in Arabic, a language renowned for its vastness and complexity [8, 9]. Muslims engage in daily recitations of the Qur'an as a religious practice aimed at drawing them closer to their Lord. In this endeavor, they adhere to the rules of Ahkam Al-Tajweed, encompassing principles such as Al-Edgham, Al-Aqlab, and Endowment [10]. Also, the Holy Qur'an stands as the most genuine and unaltered religious text throughout history, and it is the responsibility of every Muslim to safeguard its authenticity and integrity [11]. Presently, a majority of Muslims utilize the internet for online education encompassing religious principles, Qur'anic recitation and memorization, as well as for banking and social interactions [12, 13]. The Qur'anic content is widely accessible on numerous websites, presented in the form of simple text or images. The primary foundation of Islamic guidance is the Qur'an. It serves as the fundamental source in Islam, encompassing the divine words revealed to Prophet Muhammad (Peace Be upon Him). The entirety of the text has undergone thorough authentication [14]. Technological progress has facilitated the constant accessibility of both Qur'an and Hadith resources through relevant websites, enabling Muslims to explore specific topics of interest. This capability is achieved through the implementation of text categorization. Text categorization, also known as classification, involves categorizing unlabeled textual documents into predefined categories based on their content [15]. Various algorithms for text categorization have been devised, including Naïve Bayes (NB), Support Vector

Machines (SVM), k-nearest neighbors (KNN), and neural networks [16-18]. Considering all the context mentioned above, this literature survey navigates through this burgeoning field, shedding light on seminal contributions, innovative applications, and the evolving landscape of computational approaches within Qur'anic Studies and embarks on a comprehensive literature survey, aiming to illuminate the diverse applications and methodologies that have emerged in the integration of ML within the realm of Qur'anic Studies. Section 1 explains the brief introduction of the title, section 2 discussed the extensive literature survey (comprises 16 sub-sections) in great detail. Comparative study of the literatures has listed in section 3 followed by 11 meaningful observations (refer to sub-section 3.1). The last section 4 provides the conclusion and future scope of the work.

2. Literature Survey

In this section we have presented an extensive literature survey of 34+ papers in related domain. The study has based on the effective parameters such as nature of proposed work (thrust area), research gap, problem statements, tools and techniques, problem resolution, and significance of the result. Following subsequent sections describe the survey results.

Tajweed Rule Classification

The regulations governing the recitation of the Qur'an, known as Ahkam Al-Tajweed, set the correct pronunciation to be employed when reading the Holy Qur'an. Numerous existing automated systems for Qur'an recitation primarily concentrate on fundamental components, giving priority to accurate word pronunciation but neglecting other crucial elements of Ahkam Al-Tajweed. These elements encompass the rhythmic and melodious aspects of recitation, such as knowing where to pause and how to elongate or blend certain letters. Current resources addressing these latter aspects are restricted in terms of the rules covered or the sections of the Qur'an dealt with. This manuscript endeavors to bridge these gaps by tackling the challenge of identifying and applying Ahkam Al-Tajweed throughout the entire Qur'an. Our focus specifically centers on eight Ahkam Al-Tajweed encountered by novice learners of recitation. During the initial phase of our study, we extracted features using standard audio processing methods (e.g., Wavelet Packet Decomposition (WPD), Mel-Frequency Cepstral Coefficient (MFCC), Linear Predictive Code (LPC), and Hidden Markov Model-based Spectral Peak Location (HMM-SPL)). We utilized k-Nearest Neighbors (KNN), Support Vector Machines (SVM), and Random Forest (RF) classification techniques on an internal dataset consisting of thousands of audio recordings reciting every instance of the rules under consideration in the Holy Qur'an by men and women. In the first section of our research, we show that incorporating deep learning techniques can increase classification accuracy to over 97.7%.

Specifically, this accomplishment involves incorporating traditional features alongside those obtained using Convolu-

tional Deep Belief Network (CDBN), with SVM employed for classification. Our study aims to develop a system with the capability to identify specific Ahkam Al-Tajweed in a given audio recording of Qur'anic recitation. Our research focuses on eight specific Ahkam Al-Tajweed, namely "EdgamMeem" (one rule), "EkhfaaMeem" (one rule), "Ahkam Lam" in the 'Allah' term (two rules), and "Edgam Noon" (four rules). Our analysis goes beyond the mere identification of these rules and includes both accurate and incorrect applications of each rule. *This classification task consists of 16 distinct classes.* It's important to mention that our methodology is distinct from previous studies because we analyze the entire Holy Qur'an. To elaborate on our methodology, we commence by providing insights into the traits of the dataset we employed. Then, we delve into the complexities of the feature extraction and classification processes, utilizing established methods from the field of speech processing.

This comprehensive approach ensures a thorough examination of the Ahkam Al-Tajweed in various recitations, contributing to a more nuanced understanding of their usage within the context of the Qur'an [19].

Semantics Analysis of the Qur'an

This paper presents a conclusion that the application of machine learning to the interpretation of the Qur'an involves two crucial analyses. The first analysis is the semantic analysis, which is utilized for the acquisition of the dataset. The second analysis is the Natural Language processing analysis, which is employed for the categorization and grouping of data, thereby facilitating the application of machine learning algorithms. The field of interpretation has witnessed noteworthy advancements. Firstly, there is no restriction on the availability of a tool for Qur'anic interpretation that is capable of generating objective religious information without being influenced by the subjectivity of the interpreter. This tool establishes a connection to obtain information that directly refers to the meanings presented in the Qur'an. Secondly, the integration of Qur'anic data within the framework, structure, and methodology of this interpretation has a significant influence on advancing global studies in the field of interpretation development. Currently, the technological tools associated with the Qur'an mainly focus on translation and linguistic analysis. However, the extensive semantic complexity of the Qur'an opens up endless possibilities for designing interpretation analysis models, proving to be a pivotal factor in the successful implementation of machine learning algorithms. This inherent richness in semantic content enhances the depth and breadth of interpretational studies, extending beyond traditional linguistic aspects and allowing for a more nuanced exploration of the Qur'an's meanings through advanced computational approaches [20].

Qur'anic Text Processing

Establishing a framework for the analysis of Arabic text and producing statistical information that is useful to the research community is the aim of this paper. The main focus of this study is to preprocess the text of the Holy Qur'an and

apply various text mining techniques to uncover fundamental insights about its terms. The obtained results reveal several notable attributes, including significant words, a word cloud, and chapters with high occurrences of specific terms. Based on term frequencies computed with the Term Frequency-Inverse Document Frequency (TF-IDF) and Term Frequency (TF) techniques, some conclusions are drawn. Future endeavors in this field involve improving the preprocessing of the Qur'anic text by developing a more efficient and accurate algorithm, potentially capable of providing word stems similar to light stemmers. If such an algorithm is successfully developed, subsequent research will explore the extraction of knowledge and information that is beneficial to humanity through the utilization of machine learning techniques [21].

Reciter Recognition using MFCC Analysis with k-NN and ANN

This research presents a novel machine learning approach with the objective of identifying the individual who recites the Holy Qur'an. The approach involves a systematic recognition system that encompasses various key phases, such as acquiring data, pre-processing, extracting features, and classification. Ten well-known reciters who lead prayers in the holy mosques of Mecca and Madinah were chosen in order to create the dataset. Mel Frequency Cepstral Coefficients (MFCC) analysis was performed on the audio file to provide a complete analysis. Pitch was used as the feature for training the K nearest neighbor (KNN) classifier and the artificial neural network (ANN) classifier, which were both used for classification purposes. Importantly, the system achieves high levels of accuracy, with the ANN classifier achieving an accuracy of 97.62% for chapter 18 and 96.7% for chapter 36. In a similar fashion, the KNN classifier delivers an accuracy rate of 97.03% for chapter 18 and 96.08% for chapter 36. These results highlight the effectiveness of the proposed system in accurately recognizing reciters based on their Qur'anic readings. The utilization of advanced machine learning techniques, combined with pitch as a distinctive feature, contributes to the robust performance of the system in identifying and attributing Qur'anic recitations to specific reciters [22].

Reciter Style Recognition using BLSTM

Intelligent systems based on speech, employing deep learning, are gaining significance in various applications. While voice signal processing efforts have primarily concentrated on the English language, limited attention has been given to the Arabic language and the Qur'an, which serves as the central religious text of Islam. The purpose of this study is to develop a system for identifying speakers based on deep learning, which incorporates Qur'anic recitations. The utilization of Bidirectional Long Short-Term Memory (BLSTM), a type of Recurrent Neural Networks (RNNs) well-known for their aptness in speech modeling and processing, is recommended. Results indicate that the BLSTM-based Qur'anic speaker identification outperforms previous approaches and is also cost-effective. The proposed framework, employing BLSTM for Qur'an reciter identification, exhibits promising

results that significantly surpass those reported in the literature utilizing conventional Artificial Neural Networks (ANNs) and other baseline models. The BLSTM system, which operates on a two-tier structure, exhibits superior performance in comparison to the current state-of-the-art ANN solution. It achieves an impressive accuracy rate of up to 99.89%, surpassing the 91.28% accuracy obtained with 40 hidden layers. This serves to emphasize the heightened accuracy and cost-effectiveness that the proposed solution possesses. In future research endeavors, efforts will be directed towards expanding the scope of this work on RNNs in order to model various aspects of Qur'anic recitation [23].

Qur'anic Text Recognition

This study presents a succinct overview of a Qur'anic OCR system that employs a CNN followed by an RNN. The investigation involves the construction of six deep learning models to evaluate the impact of different input and output representations on the accuracy and performance of the model. The research compares models that utilize LSTM and GRU, and introduces a novel dataset for Qur'anic OCR that is based on the printed version of the Holy Qur'an (Mushaf Al-Madinah). This dataset includes images of Qur'anic pages and lines of text along with corresponding labels. The main contribution of this work lies in the development of a Qur'anic OCR model that can accurately recognize diacritic text in Qur'anic images. The comparison of Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) models in Arabic text recognition shows an enhanced Word Recognition Rate (WRR) and Character Recognition Rate (CRR) in the experimental results. Furthermore, a publicly available database specifically designed for Arabic text recognition research is established, integrating diacritics and the Uthmanic script. The system proposed attains a validation accuracy of 98%, Word Recognition Rate (WRR) of 95%, and Character Recognition Rate (CRR) of 99% on the test dataset [24].

Reciter Style Recognition using MFCC Analysis with NB and RF

The recognition of the individual who recites the Holy Qur'an with Tajweed holds significant importance within the Muslim community. In this investigation, a machine learning approach is implemented to identify Qur'an reciters. This is accomplished by utilizing a database containing recordings of twelve Qaris reciting the last ten Surahs of the Qur'an. The inclusion of twelve Qaris presents a classification challenge with twelve distinct categories. Two dissimilar methodologies are utilized to represent the audio data: one involves analyzing the audio in the frequency domain, while the other treats the audio as images by means of spectrograms. In the first approach, the model learning utilizes Mel Frequency Cepstral Coefficients (MFCC) and Pitch as features. For the second approach, Auto-correlograms are used for image representation. Feature learning involves employing classical machine learning methods like Naïve Bayes, J48, and Random Forest. Notably, the classifiers successfully discern the differences between the classes, achieving an 88% recognition accuracy

with Naïve Bayes and Random Forest when representing audio using MFCC and Pitch features. This demonstrates the effective recognition of Qaris based on the recitation of Qur'anic verses [25].

Quranic Topic Classification

In this paper, the main goal is to utilize imbalanced classification techniques such as Synthetic Minority Oversampling Technique (SMOTE), Random Over-Sampling (ROS), and Random Under-Sampling (RUS) to categorize imbalanced Qur'anic topics. Various metrics, including sensitivity/recall, specificity, overall accuracy, F-measure, G-mean, and MCC, are used to evaluate the experimental results. For the classification task, KNN, J48, voted perceptron, and LibSVM classifiers were employed. The study found SMOTE to be the most effective approach for classifying imbalanced Qur'anic topics. The results demonstrate an improved performance in Qur'anic classification through the application of imbalanced classification techniques [26].

Hadith Classifier

Differentiating between accurate ("Sahih") and inaccurate ("Da'ief") Hadiths is a pivotal task in the field of "Hadith judgment" science. This paper proposes the potential for constructing an automated classifier to categorize Hadiths utilizing Deep Learning techniques. Future work will involve exploring various Deep Learning models to establish an initial model for Hadith classification and automatic judgment, followed by a comparative analysis of these models [27].

Reciter Recognition using MFCC Analysis with SVM and ANN

The process of identifying the reader or reciter of the Holy Qur'an entails the recognition of different characteristics within the corresponding acoustic waveform. Within this research, a particular corpus has been developed, which comprises of 15 recognized individuals who have recited the Holy Qur'an. To extract features from the acoustic signal, we have utilized Mel-Frequency Cepstrum Coefficients (MFCC) and formed the reader's features matrix. For recognition purposes, we have employed Support Vector Machine (SVM) and Artificial Neural Networks (ANN). The experimental findings demonstrate that the Holy Qur'an Reader Identification System achieves an accuracy of 96.59% using SVM, while the accuracy is 86.1% when ANN is used [28].

Quranic Text Classification using FS Approach

Selecting features is an important aspect in text classification tasks, especially in preparing text data for labeling. However, current Feature Selection (FS) techniques have their drawbacks. Filter-based methods show lower accuracy, while wrapper-based techniques are computationally expensive. In this research, a two-step FS method is introduced. Initially, a chi-square (CH) filter-based technique is used to reduce the dimensionality of the feature set. This is followed by the application of the wrapper-based correlation-based technique to identify the most relevant features. The main goal of this method is to reduce computational runtime while maintaining high classification accuracy. The proposed approach has been

implemented to label examples of input information (Qur'anic verses) using classifiers such as NB, SVM, and decision trees. The outcomes suggest that the method suggested attains a 93.6% accuracy in just 4.17 seconds. The feature selection process, as examined in this study, proves to be essential in text classification tasks. Recognizing the limitations of existing FS techniques, this study proposes a hybridized FS method that combines chi-square and CFS algorithms in a two-step approach. The objective is to mechanize the categorization of Qur'anic verses with enhanced precision in classification and reduced computational time. The CH-CFS method proposed attains an aggregate precision of 93.6% in a time frame of 4.17 seconds, surpassing the CFS algorithm based on wrapping, which achieves the same level of precision but at a significantly higher computational time of 119.6 seconds. Future endeavors will concentrate on expanding the suggested hybrid CH-CFS algorithm to tackle additional classification predicaments [29].

Recitation Style Identifications

This research examines the efficiency of Holy Qur'an recitation based on established rules, with a focus on the recognition of recitation types using the SVM learning algorithm. Acoustic waves from ten Qur'anic reading styles ("Qira'ah") were collected and labeled in a corpus. MFCC properties were extracted and labeled, and the resulting matrix was utilized to train SVM using the WEKA tool. Testing on 30% of the labeled matrix confirmed the validity and reliability of the model. A comparison with other learning algorithms demonstrated the superiority of SVM, achieving an accuracy of approximately 96% and outperforming alternative methods such as ANN. Although SVM necessitates more time in the training phase, it proves to be faster in testing input data, highlighting its effectiveness in Qur'anic recitation recognition [30].

Text Classification using k-NN, SVM, NB and J48

The classification of verses in the Qur'an into predetermined categories holds considerable importance in the field of Qur'anic studies.

Recent progress in information technology and machine learning has led to the development of various algorithms designed for text classification tasks. Automated Text Classification (ATC) is a well-established machine learning technique that involves creating models capable of being trained to automatically assign known labels from a predetermined set to text instances. This study applies four traditional Machine Learning (ML) classifiers - SVM, NB, Decision Trees, and k-NN - to categorize specific verses from the Qur'an into three pre-defined class labels: faith (iman), etiquettes (akhlak) and worship (ibadah). The study makes use of a dataset containing verses from chapter two (al-Baqara) of the Holy Quran. Accuracy scores of the classifiers are over 80%, with the Naïve Bayes (NB) algorithm achieving the highest overall accuracy at 93.9% and an AUC of 0.964. Using machine learning techniques, this paper introduces an automated approach for classifying input Qur'anic verses. The research utilizes four traditional machine learning algorithms: SVM,

Naïve Bayes, J48, and k-NN. The features are derived from the textual information of the Qur'an by adopting established machine learning techniques. To preprocess the data and tackle the issue of dimensionality, the InfoGain and chi-square Feature Selection (FS) methodologies are applied. The classifiers are trained using the preprocessed text data along with label information. Throughout the experimentation process, the traditional 10-fold cross-validation method is consistently used. The classifiers consistently reach accuracy scores higher than 80%, with Naïve Bayes (NB) showing the highest accuracy result of 93.9% and an AUC value of 0.964. The research endeavors to investigate various other application domains within the realm of classification [31].

Text Categorization using NB, SVM, and k-NN

This study introduces a text categorization approach for classifying specific categories by examining the connections among resources. The chosen categories for comparison include Hajj, Prayer, and Zakat. Term weighting, specifically employing Term Frequency-Inverse Document Frequency (TF-IDF), combined with classification methods. In this investigation, three classification techniques-Naive Bayes (NB), Support Vector Machine (SVM), and K-Nearest Neighbor (KNN)-are employed in conjunction with term weighting (TF-IDF). The findings reveal that SVM proves effective in addressing the interrelation between Qur'an and Al-Hadith for both single- and multi-label classifications. Whether used independently or with term weighting (TF-IDF), the SVM method demonstrates superior accuracy compared to other methods, displaying a 10-20% enhancement in accuracy compared to alternative methods [32].

Kufic Text Recognition

This paper introduces a methodology for the classification and identification of text written in the Kufic script, which is one of the renowned scripts in the Arabic language. The existing approach, which is based on character segmentation, proves to be inadequate in accurately recognizing Kufic text due to its intricate nature. In contrast, the proposed system relies on word segmentation and utilizes the feature extraction techniques of Histogram of Oriented Gradient (HOG) and Local Binary Pattern (LBP). In addition, this paper includes a comparative evaluation that compares the numerical results of the new technique with those of previous methods for recognizing Arabic text. This is done to show the effectiveness of the current study. The outcomes of this approach demonstrate a 97.05% accuracy in recognizing Arabic text in the Kufic script, using the Polynomial kernel of the SVM classifier. According to the experimental findings, the proposed system performs better than previous recognition systems for Arabic text when it comes to recognizing the Kufic script [33].

Text Categorization using DL

This paper presents a methodology for the classification of verses from the Qur'an into multiple categories using Deep Learning (DL) algorithms, specifically Recurrent Neural Networks (RNN) and Convolutional Neural Networks (CNN). The methodology involves several stages, beginning with the

collection and organization of a labeled dataset comprising Qur'an verses. These verses are then converted into numerical sequences to be processed by the DL models. To capture the semantic aspects of the text and enhance the performance of the DL models, the skip-gram algorithm of Word 2Vec is employed. The embedding vectors acquired from the skip-gram algorithm are introduced to both the RNN and CNN models in order to conduct classification. The proposed technique offers a methodical approach to precisely categorize verses from the Qur'an based on 12 pre-defined primary subjects. The skip-gram algorithm of Word 2Vec is employed to take into account the meaning of the text and enhance the performance of the deep learning models. Furthermore, the performance of the deep learning classifiers is assessed using metrics such as accuracy, recall, precision, F1-score, and hamming loss. Cross-validation is employed to ensure to

ensure more reliable results. The best outcomes were achieved in terms of accuracy, recall, precision F1-score, and hamming loss, with values of 90.38, 92.49, 96.98, 93.81, and 0.0126, respectively.

The utilization of deep learning algorithms with a larger dataset yielded enhanced classification of Qur'an verses compared to conventional machine learning (ML) approaches [34].

3. Comparative Study and Observation

Based on the knowledge summarized in section 2 of literature survey, Table 1 represents a summarized comparative study of the literature; some useful insights drawn from the following Table has also discussed in this section.

Table 1. Summarize study of the literatures mentioned in section 2.

#	Work	Problem Statement	Tools/Techniques	Results/Problem Resolution
1.	Al-Ayyoub and Damer [19]	Tajweed rule classifications of Qur'anic Recitation	CDBN, KNN, SVM and RF	Achieves accuracy of 97.7% in Tajweed rule classifications
2.	Putra and Yusuf [20]	Semantic analysis and Natural Language Processing of Tafsir al-Qur'an	Different Machine Learning techniques	Semi-supervised Machine Learning algorithms are used to study labeled and unlabeled data, All these results are based on term frequencies that are calculated using both TF and TF-IDF methods All these results are based on term frequencies that are calculated using both TF and TF-IDF methods All these results are based on term frequencies that are calculated using both TF and TF-IDF methods
3.	Alhawat [21]	Processing the Text of the Holy Qur'an	Different text mining techniques	All All these results are based on term frequencies that are calculated using both TF TF-IDF methods All these results are based on term frequencies that are calculated using both TF and TF-IDF methods All the important results like, most important words, its wordcloud and chapters with high term frequency using both TF and TF-IDF methods.
4.	Alkhateeb [22]	Recognizing the Holy Qur'an Reciter	KNN and ANN	The ANN achieves accuracy 97.62% and KNN achieves accuracy 97.03% for chapter 18 while ANN achieves accuracy 96.7% and KNN 96.08 for chapter 36.
5.	Qayyum, Siddique and Qadir [23]	Recognizing the Holy Qur'an Reciter	ANN and BLSTM	The ANN achieves accuracy 91.28%and BLSTM achieves accuracy 99.89%.
6.	Masnizah <i>et al.</i> [24]	Arabic Optical text recognition	LSTM and GRU	The proposed system achieves a validation accuracy of 98%, WRR of 95%, and CRR of 99%.
7.	Khan, Qamar, and	Recognizing twelve Qaris recitation of the	Naïve Bayes, J48, and Random Forest	achieves 88%recognition accuracy with Naïve Bayes

#	Work	Problem Statement	Tools/Techniques	Results/Problem Resolution
	Hadwan [25]	last ten Surahs of the Qur'an		
8.	Arkok and Zaki [26]	Qur'anic topic classification based on imbalanced classification	LibSVM, KNN, J48, voted perceptron, SMOTE, ROS and RUS	SMOTE is considered as best approach for imbalanced classification.
9.	Najeeb [27]	Hadith classifier	Deep Learning techniques	Rule Embedded NN (ReNN), MLP, RNN, CNN, The Attention-based model, the transformers models and (GNNs) models are used.
10.	Nahar <i>et al.</i> [28]	Recognizing the Holy Qur'an Reciter	SVM and ANN	The SVM achieves accuracy 96.59% and ANN achieves accuracy 86.1%.
11.	Adeleke <i>et al.</i> [29]	Qur'anic text classification	CH and CFS for Feature selections and NB, SVM and J48 as classifiers	Achieved accuracy result of 93.6% at 4.17secs
12.	Nahar <i>et al.</i> [30]	Recitation style identification	SVM and other classifiers	Achieved accuracy result of 96%
13.	Adeleke <i>et al.</i> [31]	Automating Qur'anic verses classification using machine learning approach	SVM, Naïve Bayes, J48 and KNN	The Naïve Bayes (NB) classification algorithm attained an impressive overall highest accuracy rate of 93.9% and an AUC value of 0.964.
14.	Rostam and Malim [32]	Text categorization in Qur'an and Hadith	Naïve Bayes, SVM and KNN	The SVM method demonstrates superior accuracy compared to other methods.
15.	Zafar and Iqbal [33]	Text classification and identification of Kufic script	HOG, LBP and SVM classifier	Accuracy of 97.05% in recognizing Kufic script.
16.	Alashqar [34]	Classification of the Qur'an verses	RNN and CNN	The RNN model attained the highest accuracy and recall at 90.38% and 92.49% respectively, whereas the CNN model has demonstrated the highest precision and F1-Measure at 96.98% and 93.81% respectively.

Observations

Based on the summarized data in Table 1, following observations can be made:

Table 2. List of observations inferred from Table 1.

Observation 1:	Accuracy tends to 97% or above if tools and techniques are KNN or ANN [19, 22, 23, 24].
Observation 2:	Naïve Bayes, J48 doesn't perform well in comparison of KNN, ANN, and SVM [26, 28, 31, 33].
Observation 3:	Neural network-based classifiers perform well in comparison to Naïve Bayes, J48, and SVM [19, 22, 23].
Observation 4:	Reciter recognition accuracy is larger in SVM, ANN and KNN [22, 23, 30] as compared to Naïve Bayes [26].
Observation 5:	Classification is faster in CNN as compared to RNN [34].
Observation 6:	Reciter recognition accuracy is largest in BLSTM [23] as compared to all tools/techniques.
Observation 7:	Text classification using Deep Learning approach [24] is better than [28, 29, 33]
Observation 8:	Kufic text recognition using SVM with HOG and LBP is better than [29, 31]
Observation 9:	LibSVM + SMOTE archives better accuracy [26] for Qur'anic topic classification based on imbalanced classification.
Observation 10:	In Hadith classification, the Deep Learning approach has better accuracy than other approaches.

Observation 1: Accuracy tends to 97% or above if tools and techniques are KNN or ANN [19, 22, 23, 24].

Observation 11: In recitation of style identification, SVM has better accuracy than ANN [30].

4. Conclusion

The main contribution of the work is to bring forth the impact of machine learning algorithms in Quranic texts (Arabic and Kufic) recognition, recitation, recitation style, Tajweed, dielectric analysis and other related problems. Table 1 represents the outcomes of the work in context to the above-mentioned benchmarks. Based on the outcomes listed in the Table (refer to Table 1), Section 3.1 inferred some observations that contributed to the novelty of the study. These observations would be helpful in algorithm application/selection for ongoing research in the field of study. Future work may extend to exploring the impact of hybrid applications of algorithms for classification, feature selection, accuracy optimization, and feature extraction from acoustic signals by replacing the MFCC with Perceptual Linear Prediction (PLP), Discrete Wavelet Transformation (DWT), and Linear Prediction Cepstral Coefficient (LPCC) or with an amalgamation of them. Other possible future insights may gravitate towards the application of the existing work to other sacred texts like Hebrew, Vedic, Pali, Avesta, etc. No algorithm can outperform in every situation, so we need rigorous research to get an ideal benchmark.

Abbreviations

AI	Artificial Intelligence
ANN	Artificial Network
ATC	Automated Text Classification
BLSTM	Bidirectional Long Short-Term Memory
CDBN	Convolutional Deep Belief Network
CH	Chi-square
CNN	Convolutional Neural Networks
CRR	Character Recognition Rate
DL	Deep Learning
DWT	Discrete Wavelet Trans-formation
FS	Feature Selection
GRU	Gated Recurrent Unit
HMMSP	Hidden Markov Model-based Spectral Peak Location
HOG	Histogram of Oriented Gradient
KNN	k-nearest Neighbors
LBP	Local Binary Pattern
LPC	Linear Predictive Code
LPCC	Linear Prediction Cepstral Coefficient
LSTM	Long Short-Term Memory
MFCC	Mel-Frequency Cepstral Coefficient
ML	Machine Learning

NB	Naïve Bayes
PLP	Perceptual Linear Prediction
RF	Random Forest
RNNs	Recurrent Neural Networks
SVM	Support Vector Machines
TF	Term Frequency
TF-IDF	Term Frequency-Inverse Document Frequency
WPD	Wavelet Packet Decomposition
WRR	Word Recognition Rate

Declarations Ethical Approval

Not Applicable.

Funding

Not Applicable.

Conflicts of Interest

The authors declare no conflicts of interest.

References

- [1] Mahesh B. *Machine learning algorithms-a review*. International Journal of Science and Research (IJSR). 2020 Jan; 9(1): 381-6, <https://doi.org/10.21275/ART20203995>
- [2] Madadzadeh F, Bahariniya S. *The Role of Artificial Intelligence in Understanding and Interpreting the Quran*. Journal of Community Health Research. 2024 Jan 27, <http://dx.doi.org/10.18502/jchr.v12i35.14792>
- [3] Sulistio B, Ramadhan A, Abdurachman E, Zarlis M, Tri-setyarso A. *The utilization of machine learning on studying Hadith in Islam: A systematic literature review*. Education and Information Technologies. 2024 Apr; 29(5): 5381-419.
- [4] Soufan A. *Deep learning for sentiment analysis of Arabic text*. In Proceedings of the Arab WIC 6th annual international conference research track 2019 Mar 7 (pp. 1-8), <https://dx.doi.org/10.1145/3333165.3333185>
- [5] *Wikipedia. Muslims* – Wikipedia. 2022; Available from: URL: <https://en.wikipedia.org/wiki/Muslims>
- [6] Hegazi MO, Hilal A, Alhawarat M. *Fine-grained Quran dataset*. International Journal of Advanced Computer Science and Applications (IJACSA). 2015; 6(12): 262-7, <http://dx.doi.org/10.14569/IJACSA.2015.061241>

- [7] Lawrence B. *The Qur'an: a biography*. Atlantic Books Ltd; 2014 Oct 2.
- [8] *Fluent Arabic. 3 Reasons why starting to learn Arabic is difficult*. 2022; Available from: <https://www.fluentarabic.net/why-learning-arabic-is-difficult/>
- [9] Sadi AS, Anam T, Abdirazak M, Adnan AH, Khan SZ, Rahman MM, Samara G. *Applying ontological modeling on Quranic "nature" domain*. In 2016 7th International Conference on Information and Communication Systems (ICICS) 2016 Apr 5 (pp. 151-155). IEEE.
- [10] Alsmadi I, Zarour M. *Online integrity and authentication checking for Quran electronic versions*. Applied Computing and Informatics. 2017 Jan 1; 13(1): 38-46, <https://doi.org/10.1016/j.aci.2015.08.001>
- [11] Tayan O, Kabir MN, Alginahi YM. *A Hybrid Digital-Signature and Zero-Watermarking Approach for Authentication and Protection of Sensitive Electronic Documents*. The Scientific World Journal. 2014; 2014(1): 514652, <https://doi.org/10.1155/2014/514652>
- [12] Elhadj YO, E-Halagat: *An e-learning system for teaching the holy Quran*. Turkish Online Journal of Educational Technology-TOJET. 2010 Jan; 9(1): 54-61.
- [13] Muhammad A, ul Qayyum Z, Tanveer S, Martinez-Enriquez A, Syed AZ. *E-hafiz: Intelligent system to help muslims in recitation and memorization of Quran*. Life Science Journal. 2012 Oct; 9(1): 534-41.
- [14] Shafi M. *The HADITH-How it was Collected and Compiled*. Teachers Institute Lecture. 2017.
- [15] Adeleke AO, Samsudin NA, Mustapha A, Nawawi NM. *Comparative analysis of text classification algorithms for automated labelling of Quranic verses*. Int. J. Adv. Sci. Eng. Inf. Technol. 2017 Aug; 7(4): 1419, <https://doi.org/10.18517/ijaseit.7.4.2198>
- [16] Elghazel H, Aussem A, Gharroudi O, Saadaoui W. *Ensemble multi-label text categorization based on rotation forest and latent semantic indexing*. Expert Systems with Applications. 2016 Sep 15; 57: 1-1, <https://doi.org/10.1016/j.eswa.2016.03.041>
- [17] Hassanat AB, Abbadi MA, Altarawneh GA, Alhasanat AA. *Solving the problem of the K parameter in the KNN classifier using an ensemble learning approach*. arXiv preprint arXiv: 1409.0919. 2014 Sep 2, <https://doi.org/10.48550/arXiv.1409.0919>
- [18] Opitz D, Maclin R. Popular ensemble methods: *An empirical study*. Journal of artificial intelligence research. 1999 Aug 1; 11: 169-98, <https://doi.org/10.48550/arXiv.1106.0257>
- [19] Al-Ayyoub M, Damer NA, Hmeidi I. *Using deep learning for automatically determining correct application of basic quranic recitation rules*. Int. Arab J. Inf. Technol. 2018 Apr; 15(3A): 620-5.
- [20] Putra DI, Yusuf M. *Proposing machine learning of Tafsir al-Quran: In search of objectivity with semantic analysis and Natural Language Processing*. In IOP Conference Series: Materials Science and Engineering 2021 Mar 1 (Vol. 1098, No. 2, p. 022101). IOP Publishing, <https://doi.org/10.1088/1757-899X/1098/2/022101>
- [21] Alhawarat M, Hegazi M, Hilal A. *Processing the text of the Holy Quran: a text mining study*. International Journal of Advanced Computer Science and Applications. 2015 Feb; 6(2): 262-7, <https://dx.doi.org/10.14569/IJACSA.2015.060237>
- [22] Alkhateeb JH. *A machine learning approach for recognizing the Holy Quran reciter*. International Journal of Advanced Computer Science and Applications. 2020; 11(7): 268-71, <http://dx.doi.org/10.14569/IJACSA.2020.0110735>
- [23] Qayyum A, Latif S, Qadir J. *Quran reciter identification: A deep learning approach*. In 2018 7th International Conference on Computer and Communication Engineering (ICCE) 2018 Sep 19 (pp. 492-497). IEEE, <https://dx.doi.org/10.1109/ICCE.2018.8539336>
- [24] Mohd M, Qamar F, Al-Sheikh I, Salah R. *Quranic optical text recognition using deep learning models*. IEEE Access. 2021 Mar 4; 9: 38318-30, <https://dx.doi.org/10.1109/ACCESS.2021.3064019>
- [25] Khan RU, Qamar AM, Hadwan M. *Quranic reciter recognition: a machine learning approach*. Advances in Science, Technology and Engineering Systems Journal. 2019; 4(6): 173-6, <http://dx.doi.org/10.25046/aj040621>
- [26] Arkok BS, Zeki AM. *Classification of Quranic topics based on imbalanced classification*. Indones. J. Electr. Eng. Comput. Sci. 2021 May; 22(2): 678-87, <https://dx.doi.org/10.11591/ijeecs.v22.i2.pp678-687>
- [27] Najeeb MM. *Towards a deep leaning-based approach for hadith classification*. European Journal of Engineering and Technology Research. 2021 Mar 12; 6(3): 9-15, <https://dx.doi.org/10.24018/ejeng.2021.6.3.2378>
- [28] Nahar KM, Al-Shannaq M, Manasrah A, Alshorman R, Alazzam I. *A holy quran reader/reciter identification system using support vector machine*. International Journal of Machine Learning and Computing. 2019 Aug; 9(4): 458-64, <https://dx.doi.org/10.18178/ijmlc.2019.9.4.826>
- [29] Adeleke A, Samsudin NA, Othman ZA, Khalid SA. *A two-step feature selection method for quranic text classification*. Indonesian Journal of Electrical Engineering and Computer Science. 2019 Nov; 16(2): 730-6, <https://dx.doi.org/10.11591/ijeecs.v16.i2.pp730-736>
- [30] Nahar KM, Al-Khatib RM, Al-Shannaq MA, Barhoush MM. *An efficient holy Quran recitation recognizer based on SVM learning model*. Jordanian Journal of Computers and Information Technology (JJCIT). 2020 Dec 1; 6(04): 394-414, <https://dx.doi.org/10.5455/jjcit.71-1593380662>
- [31] Adeleke A, Samsudin N, Mustapha A, Khalid SA. *Automating quranic verses labeling using machine learning approach*. Indonesian Journal of Electrical Engineering and Computer Science. 2019 Nov; 16(2): 925-31, <http://doi.org/10.11591/ijeecs.v16.i2.pp925-931>

- [32] Rostam NA, Malim NH. *Text categorisation in Quran and Hadith: Overcoming the interrelation challenges using machine learning and term weighting*. Journal of King Saud University-Computer and Information Sciences. 2021 Jul 1; 33(6): 658-67, <https://dx.doi.org/10.1016/j.jksuci.2019.03.007>
- [33] Zafar A, Iqbal A. *Application of soft computing techniques in machine reading of Quranic Kufic manuscripts*. Journal of King Saud University-Computer and Information Sciences. 2022 Jun 1; 34(6): 3062-9, <https://dx.doi.org/10.1016/j.jksuci.2020.04.017>
- [34] M Alashqar A. *A Classification of Quran Verses Using Deep Learning*. International Journal of Computing and Digital Systems. 2023 Jul 22; 16(1): 1041-53, <https://dx.doi.org/10.12785/ijcds/160176>