Digital Hadith authentication: Recent advances, open challenges, and future directions

Saqib Hakak1 | Amirrudin Kamsin2 | Wazir Zada Khan3 | Abubakar Zakari4 | Muhammad Imran5 | Khadher bin Ahmad6 | Gulshan Amin Gilkar7

1Faculty of Computer Science, University of Northern British Columbia, Prince George, British Columbia, Canada
2Faculty of Computer Science, University of Malaya, Kuala Lumpur, Malaysia
3Department of Computer Science and Information System, Jazan University, Jazan, Saudi Arabia
4Department of Computer Science, Kano University of Science and Technology, Wudil, Nigeria
5College of Applied Computer Science, King Saud University, Riyadh, Saudi Arabia
6Academy of Islamic Studies, University of Malaya, Kuala Lumpur, Malaysia
7College of Computer Science, Shaqra University, Riyadh, Saudi Arabia

Abstract
The Holy Quran and Hadith are the two main sources of legislation and guidelines for Muslims to shape their lives. The daily activities, sayings, and deeds of the Holy Prophet Muhammad (PBUH) are called Hadiths. Hadiths are the optimal practical descriptions of the Holy Quran. Technological advancements in information and communication technologies (ICT) have revolutionized every field of daily life, including digitizing the Holy Quran and Hadith. Available online contents of Hadith are obtained from different sources. Thus, alterations and fabrications of fake Hadiths are feasible. Authentication of these online available Hadith contents is a complex and challenging task and a crucial area of study in Islam. Few Hadith authentication techniques and systems are proposed in the literature. In this study, we have surveyed all techniques and systems, which are proposed for Hadith authentication. Furthermore, classification, open challenges, and future research directions related to Hadith authentication are identified.

1 | INTRODUCTION

Islam is the second largest religion in the world, and its followers use the Holy Quran and Hadith as the main source of ethical guidelines and principles. The narrations, actions, deeds, and supplications of the Holy Prophet Muhammad (PBUH) are called Hadith or Sunnah. Hadiths are the optimal practical examples that the Muslims accomplish in their daily activities. The Holy Quran is the complete message from God (Allah) in forms of verses to mankind, and Hadiths are the practical demonstration and descriptions of Allah’s message through the Holy Prophet Muhammad (PBUH). For
example, in the Holy Quran, Allah has ordered mankind to worship Him (eg, prayer—Salah), whereas Holy Prophet Muhammad (PBUH) practically demonstrates and teaches the preparation and offering of prayers (ie, steps of ablutions and prayer). Furthermore, the importance of Hadith is highlighted in the Holy Quran through different verses (eg Al-Ahzaab 33(21) and An-Noor 24(54)).

The writing, compilation, and collection of Hadith were not performed during the life of the Holy Prophet Muhammad (PBUH) and his companions (ie, Sahabah—immediate follower). The collection and writing of Hadith started in the eighth and ninth centuries into different categories. The categories of Hadith include authentic (Sahih), good (Hasan), weak (da‘ief), and fabricated or fake (mawdu’) based on two main characteristics of isnad (ie, who narrated the Hadith—chain of reporters) and matn (ie, the main text of Hadith). The authentication of Hadith includes two aspects, namely, the contents of Hadith and the technological aspect; the former, as described previously, refers to the passing down of the Hadith from generation to generation for 100 years before being appropriately collected, compiled, and written. Technological advancements of information technology and telecommunication (ICT) have revolutionized every field of daily life, including digitization and online availability of the Holy Quran and Hadith. Available online contents of Hadith are obtained from different sources. Thus, altering and fabricating the actual Hadith are highly feasible. The authentication of these online available Hadith contents is a crucial and challenging task and a major field of study in Islam.1-3 A simple scenario of Hadith authentication process is shown in Figure 1.

It is very crucial to stop the dissemination of fabricated Hadith due to its vast influence among Muslims. Fabricated Hadiths can easily corrupt minds of Muslims and propagate false information with respect to sayings of Holy Prophet Muhammad (PBUH).4 The practice of spreading fabricated Hadiths started with the murder of third Caliph, that is, Hazrat Uthman Ibn Al-Affan.4 As per various research studies, the motive for spreading fabricated Hadiths were for political gains.5 However, in today’s digital era, it is very convenient to post the fabricated Hadith and poison the minds of people who are not aware of Islamic teachings. If the trend of spreading fabricated Hadith go unchecked, it might have following negative consequences:

- As it is more convenient to read Hadiths online, there is high probability that people will start ignoring authentic hard copies of Hadiths. Currently, there are numerous Hadith repositories available online and it is hard to authenticate the people responsible for managing those websites. Hence, it is very easy to feed fabricated information into the minds of young people, which might give rise to dangerous trend of abandoning authentic hard copies.
- Fabricated Hadiths is akin to fake news and might create confusion among the people with respect to the teachings of Islam. Suppose there is a particular fabricated Hadith with respect to health that encourages people to eat a particular kind of herb and so on. People with no knowledge of Hadiths might follow that action despite of its harmful effects. In addition, relying on fabricated Hadiths might encourage Muslims to follow/invent new practice with respect to

![Figure 1: Hadith authentication process](Image)
faith, law, moral, acts of worship, and so on, which is completely prohibited in Islam. The dispersion of fabricated Hadiths also affects the mindset of Muslims. There are many fabricated Hadiths that encourage Muslims to abandon this world. This mindset has already weakened the status of Muslims worldwide, which is the prime reason Muslims are lagging far behind in the area of science and technology.

- Fabricated Hadiths can be also used as a justification to fulfill one’s personal/private desires and might lead to ideological differences among Muslims. There are people who follow Hadiths blindly and do not care to verify its authenticity. Such a trend is worrisome as that particular Hadith usually becomes part of their lives and disseminates viciously through that particular circle to other groups.

Hence, the present study highlighted the basic requirements of Hadith authentications considering the importance of Hadith in Islam and the daily lives of Muslims. In addition, a comprehensive survey of recent advancements in the field of Hadith authentication was presented. Finally, this study outlined the challenges and future research directions related to Hadith authentication.

The remainder of this article is organized as follows: Section 2 discusses the requirements of Hadith authentication. Sections 3 and 4 introduce the survey on recent advancements and the challenges in Hadith authentication, respectively. Section 5 describes the future research directions, and Section 6 presents the discussion. Finally, Section 7 provides the conclusions drawn from this study.

2 | REQUIREMENTS FOR HADITH AUTHENTICATION

To determine whether the provided Hadith is authentic and originally attributed to Holy Prophet Muhammad (PBUH), Hadith scholars have set certain detailed methods on the basis of the definition of an authentic Hadith. Hadith scholars have established various rules to authenticate a Hadith; these rules include connected sanad, trustworthy narrator, reliable narrator with the ability to memorize excellently (ābit), free from irregularity (Shādh), and free from hidden detrimental flaws (‘Illeh). However, we have discussed only three main rules that can be easily implemented using ICT technologies for authenticating Hadith. The three basic requirements, as depicted in Figure 2, include connected sanad, trustworthy narrator, and matn-based approaches.

2.1 | Connected sanad

The first basic rule for determining the authenticity of Hadith is the connected sanad rule. In this rule, the chain of narrators must be connected and conveyed by a trustworthy, completely competent person until the last narrator. In addition, the chain of narration must contain neither a serious concealed flaw (‘Illeh) nor irregularity (Shādh). The two aspects used to analyze the chain of narrators include the date of death of the narrator and biography of the narrator. This information is important because it highlights the difference in time gaps and provides us rough estimation whether a particular Hadith was written during his/her life span. This method coincides with the words of Holy Prophet Muhammad (PBUH) in a popular Hadith, that is,

أَعْمَارُ أَمْتِي مَا بَيَّنَ السَّنَنَ إِلَى السَّبْعِينَ، وَأَقْلُهُمْ مِنْ يَجُوزُ ذَلِكَ

English Translation: It was narrated from Abu Hurairah that the Messenger of Allah (PBUH) said: “The ages of (the people in) my nation will be between sixty and seventy, and few of them will exceed that.” [Sunan Ibn Majah (4326), Jami at-Tirmidhi (3550)].

This Hadith suggests that the average life span of Muslims will not exceed 70 years. This method was introduced by previous Hadith scholars, who summarized the dates of death of Hadith narrators. Thus, the date of death of the narrator is important in determining the validity of a Hadith on the basis of the chain or matn aspects. The second rule is to analyze the biography of a narrator. To analyze through biographic method, the following information must be ascertained: names of narrator teachers, names of their students, and criticisms or praises to these narrators in detail.
2.2  |  Trustworthy narrator

Every narrator of the Hadith must be fair. Fair narrators must fulfill the four conditions, as shown in Figure 2, including a Muslim, at the age of puberty (baligh), wise ('aqil), and not an open sinner. Hadith narrators must have a precise and accurate memory to avoid misleading and misusing the words of the Holy Prophet Muhammad (PBUH). The narrator memorizes all the Hadith received (heard) from his teacher. Such Hadiths can be conveyed as told to the narrator. The narrator can narrate all Hadiths at any given time. The Hadith obtained from their teachers is noted on books or sheets. Each entry in the note must be verified in the following manner:

- Contains the support of the Holy Prophet Muhammad (PBUH); sahabah's narrations.
- Encompasses the support from other narratives about the same Hadith in different times/era.
- Comprises the support from other students, who have heard the Hadith from the same teacher.
- Includes the support for the narration of Hadith narrators with his friend’s narrations when receiving the Hadith.
- Contains support and comparison between memorization and the book.
- The narration of the Hadith narrators coincides with a Quranic verse.
- Free from irregularity (Shādh).
- Free from hidden detrimental flaws ('Ilalah).

2.3  |  Matn (authentication of text content)

Matn is the secondary source of the Hadith authentication by Islamic scholars. Matn mainly focuses on the content of the Hadith in comparison with the chain of narrators. The rule is simple, that is, rejecting any Hadith whose content is opposite to any content given in the Holy Quran. The optimal example related to the matn rule is a narration, in which

![Figure 2: Required components of Hadith authentication](image-url)
Hazrat Aisha (PBUH) rejected the Hadith report presented by Ibn Umar stating that “dead relative would be punished for his family’s excessive mourning.” The reason for rejection by Hazrat Aisha (PBUH) was based on the verse mentioned in the Holy Quran, that is, “no bearer of burdens bears the burdens of another (Quran 53:38).”

3 | RECENT ADVANCEMENTS IN HADITH AUTHENTICATION

Research on Hadith authentication has remained at its infancy. However, state of the art of research on security, privacy, and authentications in other various relevant fields can be found in References 6-9. In the field of Hadith authentication, the first step is to check whether its translation to other languages has been done properly or not.10 With respect to this issue, the two potential translation techniques include:

- Rule-based approaches:11-13 In rule-based approaches, the lexicons are manually defined by the expert users of the Arabic language. The rules for translation are written by experts based on which translation is done. This technique is popular among researchers and analyses semantic and syntax levels deeply. Mostly word to word translations are done using this approach.
- Statistical-based approaches:14-17 In these approaches, the algorithms are trained to translate the given input by analyzing bilingual text corpus (existing human translations). Along with the statistics of language model and bilingual corpus, suitable translations are produced. However, one of the key limitations of this approach is nonavailability of large bilingual corpus.

In terms of Hadith authentication where we assume translation is correct (from Arabic to English), the recent advancements have been categorized as follows: chain of Hadith narration (CHN) approach, neural network approach, finite-state automation approach, and classification-based approach, as shown in Figure 3. In addition, a comparative analysis of recent advances in Hadith authentication is presented in Table 1.

3.1 | CHN approach

A CHN approach is a process of identifying the chain of Hadith transmission from one narrator to another. Hadith is a report of the sayings of Holy Prophet Muhammad (PBUH). Each report was orally transmitted from one person to another until it reached the person, who recorded the report along with the chain of transmission. These chains were finally represented as a graph of narrators. Various approaches had been proposed to verify and authenticate Hadith stories and their narrators. Shukur et al18 attempted to verify the authenticity of a Hadith story and its narrators. These authors

![Hadith Authentication Approaches](image-url)
<table>
<thead>
<tr>
<th>Technique</th>
<th>Objective</th>
<th>Advantage</th>
<th>Evaluation metric</th>
<th>Datasets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shukur et al(^{18})</td>
<td>To verify Hadiths using a narrator analysis approach</td>
<td>The findings of this research showed that information visualization approach is an appropriate tool for supporting the learning of Hadith science</td>
<td>×</td>
<td>20 students of Islamic Education Program</td>
</tr>
<tr>
<td>Saloot et al(^{19})</td>
<td>To conduct a comparative analysis of existing Hadith classification approaches</td>
<td>The result of the evaluation of the classification method reveals that neural networks classify the Hadith with 94% accuracy</td>
<td>Functionality, simplicity, ( F )-Score, and accuracy</td>
<td>3150 Hadiths from the book of Sahih Al-Bukhari</td>
</tr>
<tr>
<td>Makhlouta et al(^{20})</td>
<td>To enhance automatic recognition of named entities from Arabic text using techniques that work appropriately for the Latin-based languages.</td>
<td>In all applications, their method reports high precision and recall and learns lemmas about phrases that improve results</td>
<td>×</td>
<td>Recall, precision, and ( F )-score NLP applications (Hadith application and Biblical genealogy family tree extraction)</td>
</tr>
<tr>
<td>Rasool et al(^{21})</td>
<td>To aid in using microdata and webmasters that help build a vocabulary similar to the “type hierarchy” of Schema.org</td>
<td>They observed that using microdata of HTML5 to build their custom vocabulary for referencing the Quran verses and Hadith on the WorldWide Web is advantageous</td>
<td>×</td>
<td>Quran verses and Hadith</td>
</tr>
<tr>
<td>Kamsin et al(^{22})</td>
<td>To help provide a reliable system to assist the central body and end users to access the authenticity of digital Quran applications</td>
<td>The system can validate them from semantic and linguistic aspects.</td>
<td>Accuracy</td>
<td>Quran and Hadith</td>
</tr>
<tr>
<td>Ibrahim et al(^{23})</td>
<td>×</td>
<td>The results show that the deepest level of Hadith authentication is that which involves the principles of Hadith science into the process</td>
<td>Excluding other studies that highlight the principles of Hadith science</td>
<td>×</td>
</tr>
<tr>
<td>Hassaine et al(^{24})</td>
<td>To identify the trustworthiness of Hadith narrators</td>
<td>Classification accuracy increases with the number of annotators, who agree on the authenticity of each Hadith. The method successfully extracts relevant keywords</td>
<td>×</td>
<td>Hadith</td>
</tr>
<tr>
<td>Aldhaln et al(^{25})</td>
<td>To classify Hadith into four major classes, namely, Sahih, Hasan, Da'ief, and Maudo</td>
<td>The findings show that the accurate rate of the naïve Bayes classifier has been improved through the proposed approach at 46.54%.</td>
<td>Decision trees and naïve Bayes</td>
<td>999 Hadiths from three books: Sahih Al-Bukhari, Jamiu Al-Termithi, and Silsilat Al-AHadith Al-Daiefah Wal, Al-awduah</td>
</tr>
</tbody>
</table>

(Continues)
TABLE 1 (Continued)

<table>
<thead>
<tr>
<th>Technique</th>
<th>Objective</th>
<th>Advantage</th>
<th>Evaluation metric</th>
<th>Datasets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Najeeb²⁶</td>
<td>To build an automatic Hadith isnad processing system</td>
<td>This research opens the path to building an automatic information system to classify Hadith to Sahih or Da'ifeh through an associative classification technique.</td>
<td>Associative classification</td>
<td>Six books: Sahih Al-Bukhari, Sahih Muslim, Sunan Abu Daoud, Sunan Al Termidhi, Sunan Ibn Majah, and Sunan Al Nasai</td>
</tr>
<tr>
<td>Harrag²⁷</td>
<td>To detect and extract passages or sequences of words containing relevant information from the prophetic narrations</td>
<td>The overall precision and recall are 71% and 39%, correspondingly</td>
<td>Precision and Recall</td>
<td>2602 Hadiths from Sahih Al-Bukhari</td>
</tr>
<tr>
<td>Sayoud²⁸</td>
<td>To conduct experiments of authorship classification on the Quran and Hadith to determine whether or not they have the same author</td>
<td>This research aids in obtaining interesting information on the ancient books of origins</td>
<td>×</td>
<td>Quran and Hadith</td>
</tr>
</tbody>
</table>

developed a prototype of a CHN visualizer and tested it on 20 students of the Islamic education program. The result showed that an information visualization approach is an excellent tool for supporting the learning of Hadith science.

### 3.2 Neural network approach

Neural networks are methods for performing Hadith classification and authentication. Harrag and El-Qawasmah²⁹ used an artificial neural network (ANN) approach to classifying Hadiths. These researchers adopted a three-layer feed-forward neural network with a hyperbolic tangent (tanh) activation function in the hidden layers. The neural network was trained using a back-propagation algorithm. Furthermore, a comparative study by Al-Kabi and Al-Sinjilawi³⁰ showed that the naïve Bayes classifier provides better results than other existing classifiers in Sahih Al-bukhari.

Saloot et al¹⁹ analyzed all academic journals and conference publications that use the two main methods of artificial intelligence (AI) for Hadith text, namely, Hadith classification and mining. Hadith methods and algorithms were discussed and analyzed in terms of few parameters, such as functionality, simplicity, $F$-score, and accuracy. The evaluation result showed that neural network classifies Hadith with 94% accuracy. A total of 3150 Hadiths from the book of Sahih Al-Bukhari were considered in the dataset.

### 3.3 Finite-state automaton approach

Makhlouta et al²⁰ proposed a method for extracting entities, events, and relations among them from Arabic text using a hierarchy of finite-state machines driven by morphological features, such as part of speech and gloss tags. These authors also proposed graph transformation algorithm and evaluated their method on two natural language processing applications. The result showed that, in all applications, their method obtains reports with high precision and recall and learns lemmas on phrases that improve results.

Rasool et al²¹ used microdata of HTML5 to build a custom vocabulary for referencing the Quran verses and Hadith on the World Wide Web. These researchers also implemented two modules for generating the HTML code after embedding
the metainformation of the origin of Quranic verses and Hadith using microdata. These authors also developed two modules to allow webmasters to validate their code before adding to the webpages.

Hakak et al\(^2\) conducted a survey to identify different approaches that can be used to preserve and verify the content integrity of the Quran. The authors further highlighted several existing issues and challenges in the digital Quran and Hadith authentication research area with potential solutions.

Kamsin et al\(^22\) proposed to develop and evaluate the Quran and Hadith authentication system, which could be used as a tool/mechanism for improving the digital Quran publishing laws and users’ confidence toward digital Quran applications. The system could validate them in semantic and linguistic aspects.

Ibrahim et al\(^23\) presented a framework for authenticating isnad Hadith. The proposed framework consisted of three phases. In the first phase, theoretical authentication of Hadith was performed through specific predefined criteria and principles. In the second phase, on the basis of the criteria, several features were extracted, and the authentication mechanism was applied. In the last phase, through user acceptance test and views of a Hadith expert, the Hadith was validated. However, no experimental results were provided for verifying claims and validating the proposed framework.

### 3.4 Classification-based approach

Ibrahim et al\(^23\) conducted an analysis to produce a hierarchy with different levels of related studies in computational Hadith to link with the computational authentication of isnad Hadith science. The results showed that the deepest level of Hadith authentication is that which involves the principles of Hadith science into the process.

Hassaine et al\(^24\) proposed a novel approach completely based on the content of each Hadith. These authors created a binary relation whereby the Hadith corresponds to the objects of the relation, and the words correspond to its attributes. For each category, the keywords were obtained in a hierarchical order of importance by utilizing hyperrectangular decomposition. The result showed that classification accuracy increases with the number of annotators, who agreed on the authenticity of each Hadith. Thus, the findings indicated that their method successfully extracts relevant keywords and can be further combined with other traditional methods.

Aldhaln et al\(^25\) presented a Hadith classifier, which can categorize a Hadith into four major classes, namely, Sahih, Hasan, Da’ef, and Maudo’, in accordance with the status of its isnad (narrator chain). The experiment applied the C4.5 algorithm to extract the rules of classification. The findings showed that the accuracy rate of the naïve Bayes classifier has been improved by the proposed approach at 46.54%. A total of 999 Hadiths from three books, namely, Sahih Al-Bukhari, Jamiu Al-Termithi, and Silsilat Al-AHadith Al-Daeifah Wal, Al-awdhuah, have been considered in the dataset.

Najeeb\(^26\) suggested a Hadith authentication system through associative classification technique. To differentiate between Sahih and Da’ief Hadith, the system used different classification techniques, such as decision tree, Bayesian, ANN, and support vector machine (SVM). However, no experimental results were presented to support the authenticity and validity of the proposed system.

Harrag\(^27\) proposed a finite-state transducer-based system for detecting and extracting relevant information through a sequence of words and passage from the text of Holy Prophet Muhammad (PBUH) narrations. The proposed system used simple surface information for extracting and categorizing Hadith text. It also solved text comprehension problems. The Hadith of Sahîh Al-Bukhari was used for text extraction and mining. The experimental results showed that the proposed system achieves favorable results.

Sayoud\(^28\) investigated the original Quranic verses and the statements of Holy Prophet Muhammad (PBUH) (eg, Hadith) using authorship discrimination. The main focus of the study was to check whether any similarity exists between the Quran and Hadith by verifying their authorship (ie, whether or not the Quran and Hadith are written by the same author). The experiments were performed using AI and text mining (TM)-based algorithms such as linear regression, multilayer perception, and SVM. The experiment results showed a large difference between the vocabulary of Hadith and the Holy Quran. Similarly, text segmentation verified that the Holy Quran and Hadith are from different stylistic classes. Thus, the Holy Quran is not written by the Holy Prophet Muhammad (PBUH).

### 4 OPEN CHALLENGES

Considerable work is required for authenticating digital Hadith. We have highlighted the emerging research challenges and have provided the guidelines in Table 2. Several major challenging issues include the following aspects:
### Table 2  Open challenges and guidelines

<table>
<thead>
<tr>
<th>Challenges</th>
<th>Causes</th>
<th>Guidelines</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hadith translation</td>
<td>Complex vocabulary of Arabic language and use of diacritics</td>
<td>Exploring rule and statistical approaches</td>
</tr>
<tr>
<td>Reliable digital Hadith database</td>
<td>No comprehensive database available</td>
<td>Soft copies of Hadith books; authentic database approved by standard Islamic organizations</td>
</tr>
<tr>
<td>Authentication of image format-based Hadith</td>
<td>Complicated process; nonavailability of image-based Hadith systems</td>
<td>Data mining-based image authentication techniques; securing image-based Hadith through watermarking</td>
</tr>
<tr>
<td>Authentication of mobile-based Hadith applications</td>
<td>Availability of excessive nonofficial and unauthentic mobile apps</td>
<td>Using data mining techniques to distinguish between authentic mobile apps</td>
</tr>
<tr>
<td>Fabricated Hadith authentication</td>
<td>No centralized system or database that lists Hadiths</td>
<td>Online repository for fake Hadiths</td>
</tr>
<tr>
<td>Different writing styles</td>
<td>Different writing styles in Arabic; multiple databases are required</td>
<td>Natural language processing and pattern matching approaches can address this challenge</td>
</tr>
<tr>
<td>Improving time and space complexities</td>
<td>Processing and storing all Hadith books require high resources</td>
<td>New data representation and pattern matching-based approaches for authenticating the Holy Quran can address this challenge</td>
</tr>
<tr>
<td>Authentication based on matn</td>
<td>Matn-based authentication requires verifying the contents of Hadith on the basis of the contents of the Quran</td>
<td>AI and information retrieval-based approaches can solve this challenge</td>
</tr>
<tr>
<td>Authentication of Hadith with incorrect reference number</td>
<td>Sharing of inaccurate Hadith on social media</td>
<td>Smart mechanisms for verification of text and reference numbering through data mining; machine learning-based techniques</td>
</tr>
</tbody>
</table>

### 4.1 Hadith translation

One of the major research challenges in addressing Hadith fabrication is proper translation of Hadith. There is large lexical gap while translating Hadith text (written in Arabic) to other languages such as English. For example, if we want to translate the Arabic word “wudu,” its English translation to “ablution” do not have the same equivalence meaning as per Islamic rituals of Wudu. One other example is one of the events in Islam, that is, “Fatth Makkah.” The English translator might use the words such as “conquest” (taking control of an area by force) or “occupation (taking possession of an area)” while describing this event. However, none of these two words are suitable to describe this event. The reason is Holy Prophet Muhammad (PBUH) ordered to enter the city of Makkah peacefully without any violence. Therefore, to translate Hadith into different languages is one of the complicated research challenges. The potential solution might be not to translate those words whose proper translation is not possible. Instead, proper explanation in the form of notes can be put to highlight the significance of that particular word.\(^{10}\)

### 4.2 Need of reliable digital Hadith database

A reliable database is required for the research on Hadith authentication. The existing Hadith databases are not considered reliable and cohesive. Thus, no comprehensive database contains all the necessary Hadith used in this research domain. Therefore, a reliable digital Hadith database is required for result generalization in the field of study. This challenge can be addressed by compiling soft copies of Hadith books and developing an authentic database that can be authenticated by standard Islamic organizations.
4.3 | Authentication of image format based Hadith

A large content of Hadith is available in the form of images. However, authenticating an image-based Hadith is more challenging than the unicode-based format.31 No ground truth is available for image-based Hadith authentication to date, thereby making the authentication complex. This challenge can be addressed using data mining techniques, where all popular image-based Hadiths can be collected and authenticated using image processing techniques. Finally, the authenticated images can be disseminated on the Internet by securing through watermarking techniques to preserve the integrity of verified images.

4.4 | Authentication of mobile-based Hadith applications

The problem of disseminating Hadith is not limited to web only. It is also an extensive problem in the area of mobile computing. Countless Hadith applications are available on Android and iPhone platforms. No suitable work that has been proposed can authenticate Hadith-based applications available on mobile-based platforms.32 Therefore, authenticating Hadiths (mobile-based Hadith application) is challenging given the nature of modern-day Internet and architecture of an operating system. Thus, the question of reliability and truthfulness constantly emerge. This issue can be resolved by analyzing the total number of Hadith-based applications and using data-mining and crawling techniques.

4.5 | Identification of fabricated Hadiths

The issue of fabricated Hadiths is an interesting challenge that must be addressed. Several relatively popular Hadiths are fabricated. No centralized system or database lists the Hadiths that are fabricated in the research community. Therefore, given the urgency, a website, which will list the fabricated Hadiths, is required to help new and veteran researchers to gain access. This approach will aid the research community in obtaining quick access to Hadiths that are fabricated and excluded from the database.

4.6 | Variety of Arabic calligraphic writing styles

Most soft copies of Hadiths are written in Arabic. The issue with Arabic is its writing style. Although using different writings does not alter the meaning of the Hadith, it creates an issue on the authentication. For example, in the case of the Holy Quran, an uthmani verse cannot be authenticated using a nonuthmani database. Different approaches are proposed by the researchers to resolve the issue of authenticating multistyled Quranic texts.33,34 Thus, to authenticate different Hadith writings using one common database to avoid an overhead associated in maintaining multiple databases is a tedious and complex challenge. This challenge can be addressed by exploring natural language processing and pattern matching approaches.

4.7 | Improving time and space complexities

A vast collection of Hadith books consumes considerable computing resources. The use of the existing search approaches is relatively inefficient in searching and authenticating a large corpus of Hadith collections. A considerable amount of time is required to search and authenticate a particular Hadith from thousands of Hadith collections. Different approaches, such as designing new data representation and pattern matching-based approaches, have been proposed for authenticating the Holy Quran.3,35-38 Similar approaches can be enhanced to authenticate digital Hadiths.

4.8 | Authentic Hadith with incorrect reference number

Nowadays, social media platforms, such as WhatsApp, Facebook, and Twitter, have become common platforms for sharing Hadiths. Although, most of the times, the Hadith contents are accurate, but the reference to their sources are incorrect,
thereby leading to confusion among users. For example, a particular Hadith is mentioned in Sahih Bukhari, book number 1 and Hadith number 456. If the same Hadith is shared on social media using inaccurate reference details (e.g., the number is changed from 456 to 476), then the users are misled, thereby causing the users to hesitate even the accurate Hadiths. This has resulted in various interesting research challenges in the area of digital Hadith authentication. The possible solutions to tackle this interesting challenge might involve exploring AI and data mining based approaches by verifying the source of the Hadith and text (matn) along with the reference number.

5 | FUTURE RESEARCH DIRECTIONS

The exploration of different technologies can solve these complex problems related to Hadith authentication. This section presents a few innovative research directions in this area, as shown in Figure 4.

5.1 | Securing Hadith database using blockchain technology

Blockchain technology is a promising technology to secure multimedia contents from any malicious activity.\textsuperscript{39-41} Hadith is considered the second most sacred scripture among Muslims after the sacred Holy Quran. Thus, using blockchain technology to protect Hadith content is a promising future research direction. The blockchain technology can be used to protect the Hadith database from any type of tampering by the attackers.

5.2 | Hadith mining using AI

The research in the area of AI remains at its infancy stage. The use of AI technology in the area of Hadith analytics has considerable potential. AI can be used for ijtihad, which is an Islamic law that describes the process of making a legal decision by independently interpreting the legal sources, the Quran and the Hadith/sunnah. AI can also be used to determine the authenticity of the Hadith. Numerous technologies, including machine learning and deep learning, are based on AI and can be used in the area of Hadith mining. For example, deep learning can be used to train the algorithm for the automatic verification of a particular Hadith based on initial training.
5.3 | **Authentication of Hadith applications through edge computing systems**

Edge computing is the future computing paradigm that will surpass cloud computing considering its peculiar features of minimal delay and efficiency by bringing processing close to the user.\(^{42}\) There is a huge volume of Hadith contents available through various mobile-based applications, web-based applications, and so on, which is easily accessible via Internet. Therefore, edge computing can be utilized for authenticating the contents of those Hadith applications using real-time data processing and data analytic frameworks/services.

5.4 | **Hadith data analytics**

Data on the Internet increase daily. A large content of Hadith requires digitization and verification. Most Hadith contents are available in Arabic. When the Hadith contents have been completely digitized, a massive volume of data becomes available in GB or TB. Although the existing tools related to big data analytics can be used for Hadith analytics, the natural language processing issues with respect to Arabic content may involve different research challenges. These challenges can arise due to diacritics (special symbols used for pronunciation), different styles of Arabic writing, translation, and interpretation-related issues. Thus, another future research direction is processing numerous Hadith contents.

5.5 | **Hadith tutorials using gamification**

Gamification leads the revolution toward enhancing learning skills.\(^ {43}\) It is a process of improving the courses/tutorials using gamified elements by introducing elements, such as reward system and avatars. The same concept of gamification can be implemented for Hadith-based courses. The use of gamification for Hadith content will enhance learners’ capability and motivate students to participate.

5.6 | **Watermarking and exact matching algorithms**

Watermarking seems one of the promising approaches to authenticate the Hadith. In watermarking, some data are embedded into the multimedia digital content for the purpose of authentication. Watermarking has been used in numerous applications such as authenticating medical records and even digital Quran.\(^ {44-46}\) It would be also interesting area to explore exact matching algorithms for Hadith authentication. Exact matching algorithms are used to find all occurrences of particular word or sentence from a given text. There are different categories within exact matching algorithms such as character-based approach, hash-based approach, and so on\(^ {47}\) that have been used to authenticate the documents. The hybrid approach based on both of these approaches may help in authenticating the unicode format of Hadith.

6 | **DISCUSSION**

The works in the area of digital Hadith authentication are limited, as per the analysis presented in Section 2. Most studies have proposed theoretical frameworks to classify Hadith into weak and strong categories. However, several works have not been implemented and have limited scope. In addition, the main focus has been on authenticating Hadith on the basis of the CHN (isnad rule). No work has used other Hadith authentication rules, such as trustworthy narrator- and matn-based rule. Although few studies have used the matn-based rule, these works are relatively limited without implementation details. The area of digital Hadith authentication is relatively new and complex. The complexity arises given the numerous challenges mentioned in previous sections. Although the researchers have attempted to use machine learning approaches for classifying Hadiths, hybrid approaches are still required to handle digital Hadith authentication challenges.

7 | **CONCLUSION**

Hadith authentication is a major challenge of the Muslim community that requires immediate solutions. Limited work has been conducted in the area of Hadith authentication given the complexity of this issue. In this study, we highlight the
basic requirements that must be considered while determining the authenticity of the Hadith. The basic requirements include determining the chain of narrators and reliability of the narrators. Different rules are available for determining the chain of narration, including determining the date of death and analyzing the character of the narrator. Similarly, to find the reliability of the narrator, certain basic information, such as puberty details, memorization retainability, and cleverness, is analyzed. We divide the existing literature into four categories of the CHN, neural network, finite automation, and classification-based methods. The analysis results show no comprehensive approach that has addressed the issue of Hadith authentication. Most works have focused on classifying Hadiths. On the basis of the analysis, we have discussed several immediate research challenges in the area of Hadith authentication. This study will help future researchers in analyzing the state-of-the-art in Hadith authentication domain. The researchers may focus on any particular approach, such as neural network, to present a comprehensive solution in determining the authenticity of the Hadith.

ACKNOWLEDGMENTS
This study is supported by the University of Northern British Columbia under fund 15021 ORG 4460 and University of Malaya research grant (UMRG) under project RP043A-17HNE and the Ministry of Education Malaysia (previously known as the Ministry of Higher Education Malaysia) under FGRS grant project FP003-2016 M.I. work is supported by the Deanship of Scientific Research at King Saud University through research group project number RG-1435-051.

ORCID
Abubakar Zakari https://orcid.org/0000-0001-6488-0666
Muhammad Imran https://orcid.org/0000-0002-6946-2591

REFERENCES


How to cite this article: Hakak S, Kamsin A, Zada Khan W, et al. Digital Hadith authentication: Recent advances, open challenges, and future directions. Trans Emerging Tel Tech. 2020;e3977.

https://doi.org/10.1002/ett.3977