Proceedings Paper:
Tarmom, T, Atwell, E and Alsalka, M (2022) Deep Learning vs Compression-Based vs Traditional Machine Learning Classifiers to Detect Hadith Authenticity. In: Lossio-Ventura, JA, Valverde-Rebaza, J, Díaz, E, Muñante, D, Gavidia-Calderon, C, Valejo, ADB and Alatrista-Salas, H, (eds.) Information Management and Big Data. 8th International Conference on Information Management and Big Data, SIMBig 2021, 01-03 Dec 2021, Online. Communications in Computer and Information Science, 1577. Springer, pp. 206-222. ISBN: 978-3-031-04446-5. ISSN: 1865-0929. EISSN: 1865-0937.

https://doi.org/10.1007/978-3-031-04447-2_14

Deep Learning vs Compression-Based vs Traditional Machine Learning Classifiers to Detect Hadith Authenticity

Taghreed Tarmom[0000−0002−2834−461X], Eric Atwell[0000−0001−9395−3764], and Mohammad Alsalka[0000−0003−3335−1918]

University of Leeds, Leeds LS2 9JT, UK
{sctat, E.S.Atwell, M.A.Alsalka}@leeds.ac.uk

Abstract. Due to the increasing numbers of Hadith forgeries, it has become necessary to use artificial intelligence to assist those looking for authentic Hadiths. This paper presents detailed research on ways to automatically detect Hadith authenticity in Arabic Hadith texts. It examines the utilization of deep learning-based and prediction by partial matching (PPM) compression-based classifiers, which have not been previously used in detecting Hadith authenticity. The proposed methods were compared with the most recent method used which is machine learning. In addition, there is a detailed description of the new Arabic Hadith corpus (non-authentic Hadith corpus) created for this study and the authors' experiments, which also used the Leeds University and King Saud University (LK) Hadith corpus. The experiments demonstrate that the authentication based on Isnad obtained accuracy ranging from 84% to 93%. The authentication based on Matan obtained an accuracy range of 55% to 93%, while the accuracy range for this experiment was from 55% to 85%, which means that Isnad is the most effective part of Hadith for automatically detecting authenticity. Moreover, the experiment proved that Matan can be used to judge Hadith authenticity with an accuracy of 85%. The study also showed that PPM and deep learning classifiers are effective means of automatically detecting authentic Hadith.

Keywords: Hadith authenticity · Hadith corpus · Deep learning · Arabic Natural Language Processing.

1 Introduction

Text classification is the process of classifying a set of written information into one of a number of predefined classes. Several methods can be used to accomplish this task, including prediction by partial matching (PPM) compression-based algorithms, machine learning (ML) algorithms and deep learning (DL) algorithms.

Most Arabic natural language processing (NLP) studies concentrate on Modern Standard Arabic (MSA), such as [1], [2] and [3]. However, there is a paucity of research on the classification of Classical Arabic (CA) texts, such as Hadith.

During prophet Muhammad's mission, which lasted 23 years, there was no official scribe recording his speeches, deeds, orders or his silent approvals. However, his companions (Arabic: Sahaba) memorized the prophet Muhammad's legacy and passed it on to others. From generation to generation, his legacy was transmitted in oral and/or written form [10], until Hadith scholars collected them in books.

Hadith—the second source of Islam—refers to any action, saying, order, silent approval or other aspect of the holy prophet Muhammad's life or legacy that was delivered through a chain of narrators. Each Hadith has an Isnad—the chain of narrators—and a Matan—the action of the prophet Muhammad. Figure 1 is an example of a Hadith written in CA, which is entirely different from MSA in both vocabulary and spelling.

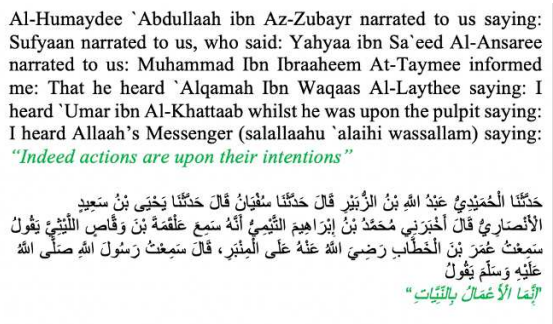

Fig. 1. An example of Hadith, Isnad in black and Matan in green.

Hadith scholars are interested in studying the validity of Hadiths because they are important in every aspect of Muslim life. In contrast to the Quran, some Hadiths, which have been handed down over centuries, have been corrupted by incompetent narrators who incorrectly transferred them. Hadith scholars classify these texts as non-authentic Hadiths.

However, many forged Hadiths have also been circulated not only by incompetent Muslims but also by pious Muslims to encourage others to follow the religious and ethical advice contained in the Hadith. The Isnad exists to clarify the Hadith's reliability. Today, however, most Muslim scholars cite Hadiths without citing their Isnad, which is contrary to the early Islamic period, when Hadiths were not cited without mentioning their Isnad [10]. Today, increased access to the internet has expanded the threat by causing a steady climb in the numbers of forged Hadiths.

To determine the authenticity of a specific Hadith, Hadith scholars use both the Isnad and the Matan. In Isnad, the narrators must be connected, so scholars study the status of each narrator to ascertain whether they are reliably accurate and connected or not [4]. Figure 2 shows an example of a Hadith determined to be non-authentic due to weak and lying narrators (highlighted in yellow).

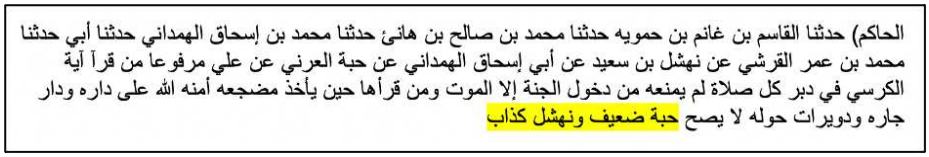
الحاكم) حدثنا القاسم بن غانم بن حمويه حدثنا محمد بن صالح بن هانئ حدثنا محمد بن إسحاق الهمداني حدثنا أبي حدثنا محمد بن عمر القرشي عن نهشل بن سعيد عن أبي إسحاق الهمداني عن حبة العرني عن علي مرفوعا من قرآ آية الكرسي في دبر كل صلاة لم يمنعه من دخول الجنة إلا الموت ومن قرأها حين يأخذ مضجعه أمنه الله على داره ودار جاره ودويرات حوله لا يصح حبة ضعيف ونهشل كذاب

Fig. 2. An example of a non-authentic Hadith due to weak and lying narrators.

In Matan, Hadith scholars study whether it is in agreement with or contradictory to Arabic grammar, another authentic Hadith or what is mentioned in the Quran. Figure 3 illustrates a Hadith that was declared non-authentic for these reasons (highlighted in yellow). In other cases, the Matan contains unacceptable words or expressions that do not reflect the Prophet Muhammad's speech or Muslim beliefs. Figure 4 shows an example of a Matan that contains an unacceptable explanation of Allah.

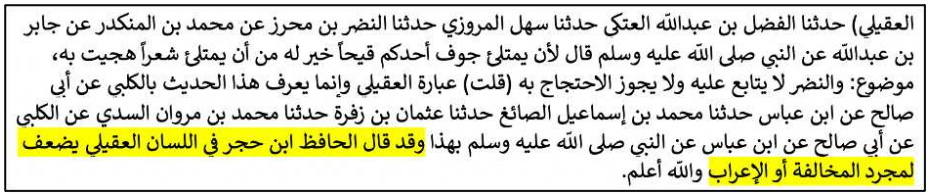
العقيلي) حدثنا الفضل بن عبدالله العتكي حدثنا سهل المروزي حدثنا النضر بن محرز عن محمد بن المنكدر عن جابر بن عبدالله عن النبي صلى الله عليه وسلم قال لأن يمتلئ جوف أحدكم قيحاً خير له من أن يمتلئ شعراً هجيت به، موضوع: والنضر لا يتابع عليه ولا يجوز الاحتجاج به (قلت) عبارة العقيلي وإنما يعرف هذا الحديث بالكلبي عن أبي صالح عن ابن عباس حدثنا محمد بن إسماعيل الصائغ حدثنا عثمان بن زفرة حدثنا محمد بن مروان السدي عن الكلبي عن أبي صالح عن ابن عباس عن النبي صلى الله عليه وسلم بهذا وقد قال الحافظ ابن حجر في اللسان العقيلي يضعف لمجرد المخالفة أو الإعراب والله أعلم.

Fig. 3. An example of a non-authentic Hadith because of the contradiction or the grammar.

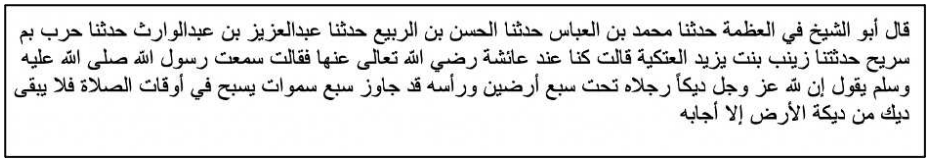
قال أبو الشيخ في العظمة حدثنا محمد بن العباس حدثنا الحسن بن الربيع حدثنا عبدالعزيز بن عبدالوارث حدثنا حرب بم سريح حدثتنا زينب بنت يزيد العتكية قالت كنا عند عائشة رضي الله تعالى عنها فقالت سمعت رسول الله صلى الله عليه وسلم يقول إن لله عز وجل ديكاً رجلاه تحت سبع أرضين ورأسه قد جاوز سبع سموات يسبح في أوقات الصلاة فلا يبقى ديك من ديكة الأرض إلا أجابه

Fig. 4. An example of a Matan that contains an unacceptable explanation of Allah.

The process of distinguishing authentic Hadiths from non-authentic Hadiths is the task of Hadith judgement science. The researchers in [5] reported that Hadiths can be automatically judged using a computerized classifier, such as ML and DL classifiers, which can assist Hadith researchers in their task.

In this paper, we first aimed to produce a new free resource for Hadith research. This language resource is called the non-authentic Hadith corpus, and it contains text from lesser-known Hadith books. Our second aim was to identify

which part of the Hadith (Isnad, Matan or both) is most effective for automatically detecting authenticity. Third, we aimed to examine the utilization of DL and PPM classifiers, which have not previously been used for detecting Hadith authenticity. The fourth aim was to compare the DL, ML and PPM classifiers to determine which is the most effective classifier when detecting the authenticity of a Hadith.

2 Related work

Hadith authentication refers to the classification of Hadiths as authentic or non-authentic using artificial intelligence methods. Several types of NLP methods can be applied to solve the problem of Hadith authentication. However, very few studies have been published in this area. In this section, we present some of the prior studies that focused on determining Hadith authenticity.

Ghazizadeh et al. [6] pointed out that determining a Hadith's authenticity involves two parameters: (1) the reliability and honesty of the Hadith narrators and (2) whether the Hadith was continuous or discrete, as determined by the Isnad. They built a fuzzy rule-based system with these parameters and expert opinions that relied on two inference engines. In the first engine, each narrator was ranked according to reliability and honesty. The output from the first step was used as the input for the second engine. This second stage produces a Hadith validation rate. To test their system, they used the *Kafi* database, which is a reliable book of Hadiths, and achieved an accuracy rate of 94%.

Bilal and Mohsin [7] noted that classifying Hadiths is a sensitive and complex task that can only be accomplished by Hadith scholars with intimate knowledge of the large number of rules involved in the process. As a result, the *Muhadith* system was built to facilitate the Hadith classification process. The aims of the *Muhadith* are to automatically classify Hadiths by imitating Hadith scholars' ability to determine authenticity. It was designed by combining ideas from distributed computing systems, web technologies and Hadith scholars' knowledge. A user types a Hadith into a web-based interface, where the Hadith then passes to the web server, where the user's input is analysed and the required data extracted. This information is then sent to the fact extractor connected to the database, which returns the results and an explanation of the Hadith classification to the user.

In their research, Aldhaln et al. [8]] used the decision tree (DT) algorithm to classify Hadiths according to degree (Sahih, Hasan, Daeef or Maudo). Their corpus consisted of 999 Hadiths from three different Hadith books: *Bukhari*, *Jami'u Al-Termithi* and *Silsilat Al-Ahadith Al-Dae'ifah w' Al-Mawdhu'ah.* . In addition, it included both the Hadiths and their attributes, as included in the Hadith books, as a means of describing their individual degrees. However, some of the Hadiths did not clearly describe these attributes, which resulted in missing values. To solve this problem, the researchers used a missing data detector (MDD). The corpus was divided into two data sets, with 66.7% of the Hadiths comprising the training data set and 33.3% comprising the test data set. Their

experiments showed that the MDD had a significant effect on the performance of the DT classifier, with accuracy rising from a rate of 50.1502% to 97.597%.

Najiyah et al. [9] asserted that non-authentic Hadiths can lead to misunderstandings of Islamic law, identifying a need to develop automatic classifications of authentic and non-authentic Hadiths. They classified Hadiths using expert systems and a DT classifier. First, they created an expert table of Hadiths by interviewing Hadith experts and confirming their findings using a variety of trusted Hadith books. They divided the Hadiths by degree into three groups: (1) Sahih, or authentic Hadiths with continuous, trustworthy Isnads and with Matan that did not contradict other authentic Hadiths; (2) Daeef , or Hadiths made weak by non-continuous Isnads, which they then divided into 17 sub-degrees; and (3) Maudo, or fabricated Hadiths created by inauthentic narrators. The degree of a Hadith can be determined by evaluating the Isnad and Matan, as authenticated by Hadith scholars. To evaluate their system, they built a training corpus containing 274 Hadiths and a test corpus containing 72 Hadiths. Their results showed that their classification model could be relied upon to classify Sahih Hadith with an error rate of only 0.00134%.

It is clear from this review of the existing research that, first, most Hadith authentication studies have focused on Isnad for automatically judging the authenticity of Hadiths, and a paper by Hakak et al. [4] indicated that authentication based on Matan is one of the challenges facing the authenticating of digital Hadith. Second, none of the existing work examines the use of DL or PPM classifiers to automatically detect a Hadith's authenticity. Thus, the present study aimed to fill these research gaps.

3 Proposed Data Sets

For this study, we selected two different Hadith corpora: a non-authentic Hadith (NAH) corpus [14] and the Leeds University and King Saud University (LK) Hadith corpus [15]. The main advantages of these corpora are that they are freely available to the Hadith research community and they have different Hadith structures.

3.1 Non-authentic Hadith (NAH) corpus

Most NLP studies for Arabic Hadith research focus on the six canonical Hadith books, and there is a shortage of research into lesser-known Hadith books. Therefore, a corpus containing Arabic Hadith from lesser-known Hadith books, in particular, would provide a new resource for Hadith community research. Thus, the purpose of the NAH corpus was to build a corpus that contained text from lesser-known Hadith books. These books are considered challenging for many reasons, including being written in a very old style, lacking a clear structure and lacking any new revisions, restructuring or editing processes.

The NAH corpus is so named due to the large number of non-authentic Hadiths it contains, as compared to authentic Hadiths. The main features of

this corpus are that it is freely available to the Hadith research community [1] and it contains 1,621,423 words from 15 non-famous Hadith books. Over 4,000 Hadiths were annotated manually according to the Hadith's *Isnad* and *Matan* in addition to *Authors comment*, *Hadith type*, *Hadith authenticity* and *Hadith topic*. These Hadiths were divided into over 7,000 Hadith records. Some of the Hadiths are classified as Hadith block, which refers to a complex kind of Hadith that contains several Isnads, Matans or author comments, and these were written sequentially. Figure 5 illustrates an example of a Hadith block. So, Hadith in Figure 5 was divided into two records since it has two isnads and matans.

<Isnad1>حدثنا عمر له أن ابن موسى حدثنا موسى بن السندي حدثنا عثمان بن
عبدالرحمن الطرايفي حدثنا عمر بن موسى بن دحية عن القاسم عن أبي أمامة
أن الله إذا غضب أنزل الوحي بالعربية وإذا رضي أنزل</Isnad1><Matan1>مرفوعاً
قال ابن حبان هذا الحديث باطل لا</Matan1><AuthorComment1>الوحي بالفارسية
أخبرني عن محمد</AuthorComment1><Isnad2>أصل له عمر بن موسى بن دحية وضاع
بن الحسين بن فنجويه حدثنا أبي حدثنا محمد بن إبراهيم حدثنا محمد بن أحمد
التميمي حدثنا أبو عصمة عاصم بن عبيد الله البلخي حدثنا إسمعيل بن زياد عن
أبغض الكلام</Isnad2><Matan2>الغالب القطان عن المقبري عن أبي هريرة رفعه
إلى الله تعالى بالفارسية وكلام الشيطان الخوزية وكلام أهل النار البخارية وكلام
قال ابن حبان وضعه إسمعيل شيخ</Matan2><AuthorComment2>أهل الجنة العربية
دجال لا يحل ذكره في الكتب إلا على سبيل القدح فيه رواه عن عاصم بن عبدالله
البلخي وهو موضوع لا أصل له من كلام رسول الله صلى الله عليه وسلم ولا حدث به أبو
هريرة ولا المقبري ولا غالب</AuthorComment2>

Fig. 5. An example of a Hadith block extracted from N3_1.

Table 1. The NAH corpus contents.

No.	Book Reference Name	Book's Title	Author	Book's Contents	Hadith's Type	No. of words	Annoteded	verifed
1	N1	الأباطيل والمناكير والصحاح والمشاهير	أبو عبد الله الهمذاني الجورقاني	Isnad/Matan/Comments	authentic and NAH	121,080	Yes	Yes
2	N2	مائة حديث ضعيف وموضوع منتشرة بين الخطباء والوعاظ	إحسان العتيبي	Matan/Comments	NAH	2,898	Yes	No
3	N3_1	اللآلئ المصنوعة في الأحاديث الموضوعة الجزء الأول ط دار المعرفة	جلال الدين السيوطي	Isnad/Matan/Comments	authentic and NAH	15,421	Yes	Yes
4	N3_2	اللآلئ المصنوعة في الأحاديث الموضوعة الجزء الثاني ط دار المعرفة	جلال الدين السيوطي	Isnad/Matan/Comments	authentic and NAH	151,382	Yes	Yes
5	N4	الأحاديث الضعيفة في كتاب رياض الصالحين	إحسان العتيبي	Isnad/Matan/Comments	NAH	5,675	Yes	No
6	N5	الجد الحثيث في بيان ما ليس بحديث ابو زيد دار الراية	أحمد بن عبد الكريم العامري	Matan/Comments	NAH	16,382	Yes	No
7	N6	الفوائد المجموعة في الأحاديث الموضوعة ط العلمية	الإمام محمد بن علي الشوكاني	Matan/Comments	NAH	139,786	Yes	Yes
8	N7	معرفة التذكرة في الأحاديث الموضوعة مؤسسة الكتب الثقافية	لابن طاهر المقدسي	Matan/Comments	NAH	115,672	No	No
9	N8	جامع الأحاديث القدسية (الضعيفة) دار الريان للتراث	عصام الدين الصابطي	Matan/Comments	NAH	246,141	No	No
10	N9	ضعيف سنن الترمذي	محمد ناصر الألباني	Isnad/Matan/Comments	NAH	663,783	No	No
11	N10	الموضوعات دار المأمون للتراث – دمشق	الحسن بن محمد الصغاني	Matan/Comments	NAH	13,508	No	No
12	N11	النخبة البية في الأحاديث المكذوبة على خير البرية المكتب الإسلامي	محمد الأمير الكبير المالكي	Matan/Comments	NAH	13,508	No	No
13	N12	المصنوع في معرفة الحديث الموضوع	علي القاري الهروي المكي	Matan/Comments	NAH	33,037	No	No
14	N13	أحاديث الإحياء التي لا أصل لها	الإمام تاج الدين السبكي	Matan	NAH	55,917	No	No
15	N14	اللؤلؤ الموضوع فيما لا أصل له أو بأصله موضوع ط دار البشائر + ط قديمة	الشيخ أبي المحاسن القاوقجي	Matan/Comments	NAH	27,233	No	No
Total	15					1,621,423		

Methodology The Web as a corpus method [11], was used to collect Hadiths from the islamweb.net and almeshkat.net websites. Because the Web texts are

[1] https://github.com/TaghreedT/NAH-Corpus.

free and written by a wide variety of writers, there is a lack of interest in proofreading [11]. We found numerous mistakes in our corpus, such as missing spaces رضياللهعنهاعنمسروق missing letters الل, and spelling errors مفوعاً, اتنهى, مرضوع. By comparing the *N3_1* Word file with the original book PDF file, we found some Hadiths missing from the Word file, and we left these errors as they were written in the source

78- الورد الأبيض خلق من عرقِ ليلة المعراج والورد الأحمر خلق من عرق
جبريلوالورد الأصفر من عرق البراق وأورده ابن فارس عن عائشة

Fig. 6. An example of missing space in N5 (in bold).

Corpus Annotation The NAH corpus contains two primary folders. The annotated folder contains seven comma-separated value (CSV) files encompassing the Hadith books that have been manually annotated for this study. The unannotated folder contains five CSV files that contain the Hadith books that have not been annotated (see Figure 7). Every Hadith in the first folder has eight primary features or attributes. These are *No.*, *Full Hadith*, the *Isnad*, the *Matan*, the *Authors Comments*, the *Hadith Type*, *Authenticity* and *Topic*. The Authenticity feature is an important label in this study and the annotator copied the Hadith authenticity from Hadith book which was acknowledged by Hadith scholars. A description of the NAH corpus features is provided in Table 2.

Table 2. Features of the NAH corpus.

Features	Description
No.	The Hadith reference number
Full Hadith	The Hadith as it appears in the book without annotations
Isnad	The chain of narrators
Matan	The act of the Prophet Muhammad
Authors Comments	The author describes the authenticity of each Hadith
Hadith Type	The Hadith Type (Maqtu' مقطوع, Mawquf موقوف and Marfo مرفوع) or Hadith degree (ضعيف, موضوع and so on)
Authenticity	Whether this Hadith is authentic or non-authentic
Topic	The chapter title

Corpus Evaluation This section describes the various experimental analyses conducted to evaluate the corpus. First, cross-corpus evaluation was used to compare the classification results of the NAH corpus with other Hadith corpora using different DL classifiers. This assisted in verifying Hadith components (Isnad and Matan) by comparing them against existing Hadith corpora. Second,

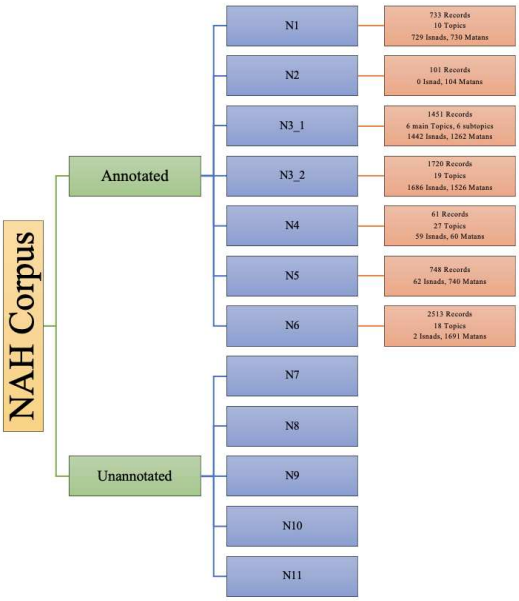

Fig. 7. The NAH corpus structure.

to verify the quality of the annotation, we applied an inter-annotator agreement (IAA) analysis.

Cross-Corpus Evaluation In order to evaluate the NAH corpus, we compared it with another existing corpus with similar features, the LK corpus [15]. In this experiment, we used one corpus as a training set and the other corpus as a testing set. The experiment used the convolutional neural network (CNN) and long short-term memory (LSTM) classifiers.

Table 3 shows that the NAH corpus identified 98% and 99% of the LK corpus using CNN and LSTM, respectively. The LK corpus identified 90% and 98% of the NAH corpus using CNN and LSTM, respectively. This demonstrates that even when using different classifiers, training models with the NAH corpus results in higher accuracy rates than training with the LK corpus.

Table 3. Cross-corpus evaluation using CNN and LSTM trained on the training datasets (rows) and tested on testing datasets (columns).

Classifier	Dataset	NAH	LK
CNN	NAH	-	98.39%
	LK	90.54%	-
LSTM	NAH	-	99.57%
	LK	98.06%	-

Inter-annotator Agreements Annotation of the NAH corpus was carried out by two annotators with Arabic and Islamic backgrounds.To validate the quality of their annotation, the Kappa coefficient, *K*, [12] was chosen to calculate the IAA between the two annotators. This paid process is quite expensive, so we provided only three books to the annotators who took part in this research. These were *N3_1*, *N3_2* and *N6* from the NAH. Then the Kappa coefficient was calculated for the total of 4,338 Hadith records and obtained Kappa values between 0.9842 and 0.9983, which indicates perfect agreement, according to [13].

3.2 Leeds University and King Saud University (LK) Hadith corpus

The LK corpus is a parallel corpus of English–Arabic Hadith built by [15], which containing 39,038 annotated Hadiths from the six canonical Hadith books. The main advantage of this corpus is that it is freely available [2] to the Hadith research community, while the main disadvantage is that the split into Isnad and Matan was automatically annotated and has only been manually verified for the *Bukhari* sub-corpus. This means that the other sub-corpora, such as Muslim, are noisy and need to be verified.

4 Deep Learning Classifiers

In this study, which used the CNN and LSTM basic models, we propose a hybrid model, the CNN–LSTM hybrid, that incorporates the advantages of each. The CNN captures the local features of the text, but with long sequences of words, it cannot preserve long-term dependencies. The LSTM overcomes the vanishing gradient problem by capturing any long-term dependencies in a lengthy sequence of words [16, 17]. Zhang et al [16] reported that this hybrid model enhanced the accuracy rate of text classification.

Word embeddings are standard representation of word meanings used in NLP [20]. Our DL models consist of an input embedding layer, a hidden layer and a dense output layer. The embedding layer is important for DL models because it permits capturing relationships between words that are hard to capture otherwise. In this layer, each word in the input data is represented by a dense vector of fixed size. We used this layer to learn an embedding for all of the words in our training datasets.

The dense output layer takes the number of classes as its output dimension. Because this is a binary classification problem, the sigmoid function was used for activation.

The hierarchy of our DL models is as follows:

- CNN. The architecture of our CNN model consists of one CNN layer with 15 filters and a kernel size of 3, followed by global max-pooling with default values (see Figure 8).

[2] https://github.com/ShathaTm/LK-Hadith-Corpus.

– LSTM. The architecture of our LSTM model consists of one LSTM layer with `hidden_nodes` = 15 and `return_sequences` = true. The `return_sequences` argument returns all the outputs of the hidden states of each time steps. The next layer is global max-pooling with default values (see Figure 8).

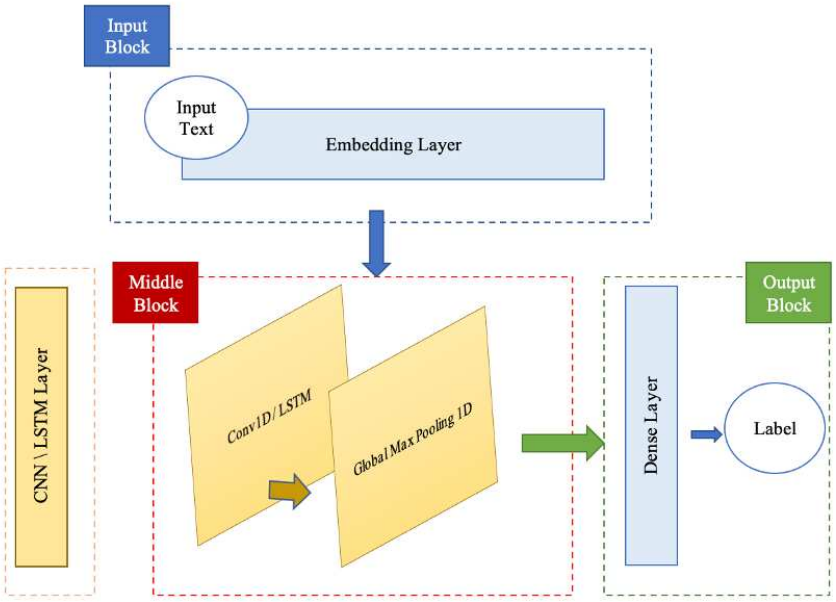

Fig. 8. The architecture of our CNN and LSTM models.

– CNN–LSTM. The architecture of this model consists of one CNN, one LSTM and global max-pooling with default values (see Figure 9).

When fitting DL models, we used a `callbacks` function with the `early_stop` method to monitor our model's performance. This method halts the training process if accuracy stops improving. In addition, we added a `patience` argument with four epochs to delay this early stopping for a set number of unimproved epochs.

5 Experiments and Results

Three experiments were performed to evaluate the automatic detection of Hadith authenticity by the compression-based classifier [18]; ML classifiers, such as the support vector machine (SVM), naïve Bayes (NB) and DT classifiers; and DL classifiers, which includes the LSTM, CNN and CNN–LSTM classifiers. These experiments were conducted to determine (1) Hadith authenticity based

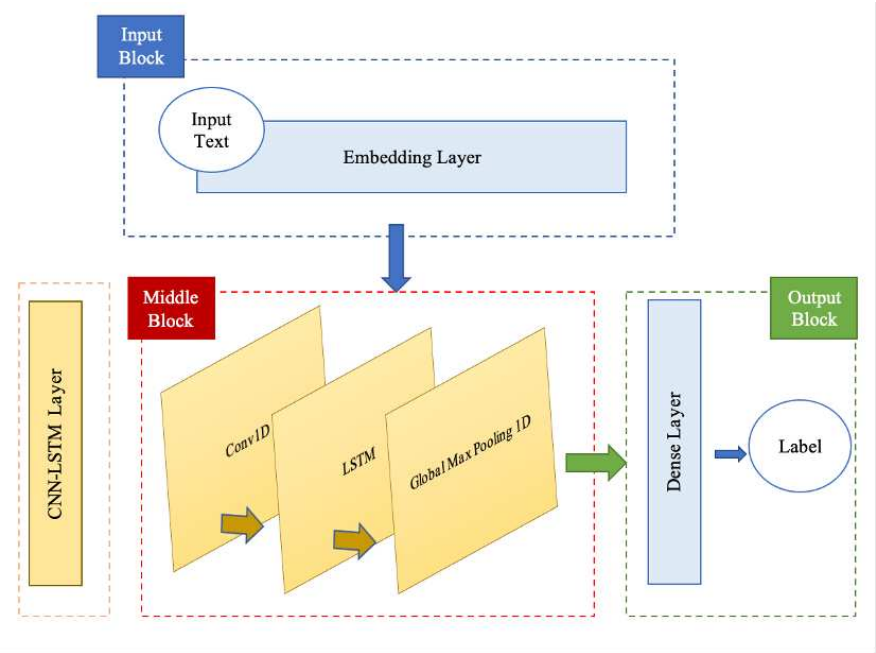

Fig. 9. The architecture of our CNN–LSTM model.

on Hadith; (2) detect Hadith authenticity based on Isnad and (3) detect Hadith authenticity based on Matan. Aside from comparing the DL, ML and PPM classifiers, the primary aim of these experiments was to identify the part of a Hadith (Isnad, Matan or both) that is best used for effective automatic determination of authenticity.

These experiments used two data sets (*N3_1* and *N3_2*) from the NAH corpus and the *Bukhari* and *Muslim* data sets from the LK corpus, with the goal of determining authenticity. Each Hadith record was assigned to one of the following two classes:

- **Authentic** records that contained an authentic Hadith from the *Bukhari* or *Muslim* data sets.
- **Non-authentic** records that contained non-authentic Hadith from the *N3_1* or *N3_2* data sets.

To guarantee a balanced distribution in the classes, we limited the number of records per class, thereby ensuring that the training phases used an equal number of records per category. We used 1,264 Hadith records from *Bukhari* to train the authentic model, and we used 1,264 Hadith records from *N3_1* to train the non-authentic model. The testing file contained 2,996 Hadith records: 1,498 Hadith records from *N3_2* and 1,498 Hadith records from *Muslim*.

5.1 Authentication Based on Hadith

In this experiment, we extracted full Hadith records containing both Isnad and Matan from the *Bukhari* and *N3_1* data sets to train the authentic and non-authentic models, respectively. The PPM and CNN–LSTM classifiers achieved higher rates of accuracy than the other classifiers by reaching up to 93%. The LSTM classifier obtained 80% and the CNN 72%. The lowest accuracy reported was from the DT classifier, with 55%. Table 4 shows the results of this experiment.

Table 4. The results of the authentication in the Hadith-based experiment.

Classifier	Accuracy (%)	Recall	Precision	F-measure
PPM	**93**	**0.94**	**0.93**	**0.93**
SVM	61	0.61	0.78	0.54
NB	76	0.76	0.83	0.75
DT	55	0.55	0.76	0.44
LSTM	80	0.97	0.73	0.84
CNN	72	0.99	0.64	0.78
CNN-LSTM	**93**	**0.93**	**0.93**	**0.93**

Figure 10 illustrates some authentic Hadiths that were predicted to be non-authentic. This might be because the narrator أبو عقيل (highlighted in blue in Figure 10) had been mentioned several times in *N3_1*, which was the non-authentic training set.

5.2 Authentication Based on Isnad

In this experiment, we extracted records that contained only Isnad from the *Bukhari* and *N3_1* data sets to train the authentic and non-authentic models, respectively. The CNN classifier achieved better accuracy than the other classifiers and reached up to 93%. This was followed by the PPM classifier, with 92%, and then the SVM classifier, with 91%. The lowest accuracy was reported for the CNN–LSTM classifier, with 84%. Table 5 presents the results of this experiment.

Table 5. The results of the authentication based on Isnad experiment.

Classifier	Accuracy (%)	Recall	Precision	F-measure
PPM	92	0.93	0.92	0.93
SVM	91	0.91	0.92	0.92
NB	89	0.89	0.90	0.89
DT	90	0.90	0.90	0.90
LSTM	90	0.95	0.87	0.90
CNN	**93**	**0.97**	**0.90**	**0.93**
CNN-LSTM	84	0.97	0.77	0.86

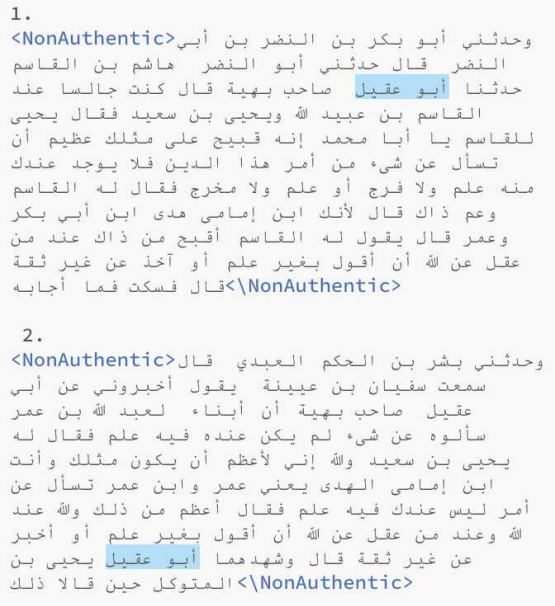

Fig. 10. Sample of authentic Hadiths predicted to be non-authentic in the PPM output of the first experiment.

Figure 11 illustrates an example of a non-authentic Isnad from *N3_2* that was predicted to be authentic. This is because the Isnad for this Hadith had a narrator, ابن عمر who is known to be a trustworthy narrator. It is possible that this Hadith was classified as authentic not because of any weakness in the Isnad but because of its Matan.

Fig. 11. Example of a non-authentic Isnad from N3_2 predicted to be authentic in the PPM output of the second experiment.

5.3 Authentication Based on Matan

In this experiment, we extracted Matan records, which contained only Matan, from the *Bukhari* and *N3_1* data sets to train the authentic and the non-authentic models, respectively. The LSTM classifier achieved the highest rates of accuracy, reaching 85%, which is lower than the previous experiments. This was

followed by the CNN and CNN–LSTM classifiers, with 84% and 82%, respectively. The PPM classifier obtained an accuracy of 79%. The lowest accuracy was reported by the SVM and DT classifiers, each with 55%. Table 6 provides the results of this experiment.

Table 6. The results of the authentication based on Matan experiment.

Classifier	Accuracy (%)	Recall	Precision	F-measure
PPM	79	0.79	0.79	0.79
SVM	55	0.56	0.75	0.45
NB	72	0.72	0.80	0.70
DT	55	0.56	0.75	0.45
LSTM	**85**	**0.80**	**0.90**	**0.85**
CNN	84	0.85	0.84	0.84
CNN-LSTM	82	0.87	0.79	0.83

Figure 12 illustrates an example of a non-authentic Matan from *N3_2* that was predicted to be authentic. This Matan was mentioned in the *Bukhari* data set several times. Furthermore, this Hadith might be narrated by different Isnad, and the Isnad mentioned in the *N3_2* data set constitutes a weakness.

<Authentic>بين كل أذانين صلاة<\Authentic>

Fig. 12. Example of a non-authentic Matan from N3_2 predicted to be authentic in the PPM output of the third experiment.

The accuracy for the first experiment ranged from 55% to 93%. The accuracy for the second experiment was between 84% and 93%, while the accuracy for this third experiment ranged from 55% to 85%, which means that Isnad was the part of a Hadith that resulted in the most effective automatic determinations of authenticity. However, this experiment also proved that we could use the Matan to judge Hadiths with an accuracy rate of 85%. Figure 13 compares the performance of PPM classifier, three ML classifiers and three DL classifiers.

6 Conclusion

This paper discussed our creation of a new Arabic corpus that uses the NAH containing samples of Arabic Hadith text from lesser-known Hadith books. Our experiments showed, first, that Isnad is the part of a Hadith that results in the most effective automatic determination of authenticity. Also proved was

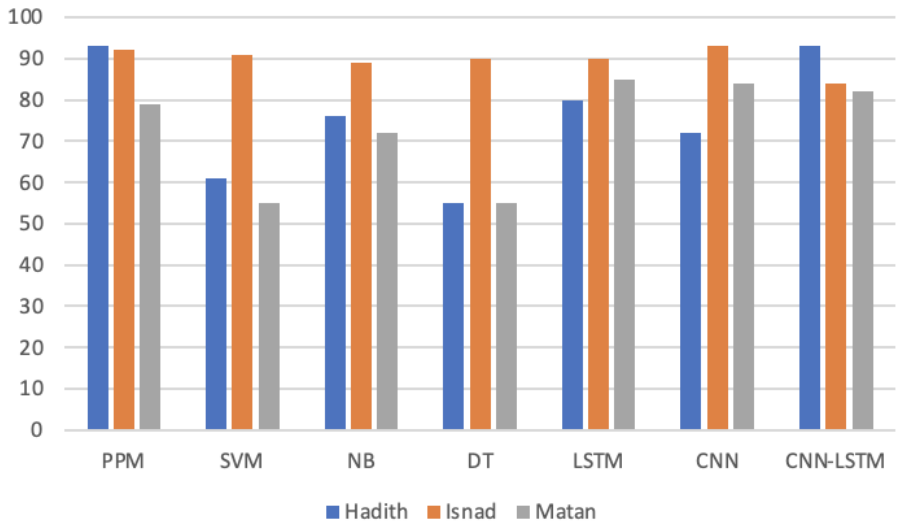

Fig. 13. A comparison of the accuracy rates using different parts of a Hadith.

that Matan can be used to judge Hadiths with an accuracy rate of up to 85%. Finally, we also demonstrated that the PPM and DL classifiers were also helpful in obtaining an effective automatic determination of Hadith authenticity.

References

1. Khreisat, L.: A machine learning approach for Arabic text classification using *N*-gram frequency statistics. J. Informetr. **3**(1), 72–77 (2009). https://doi.org/10.1016/j.joi.2008.11.005
2. Duwairi, R., Al-Refai, M.N., Khasawneh, N.: Feature reduction techniques for Arabic text categorization. J. Am. Soc. Inf. Sci. Technol. **60**(11), 2347–2352 (2009). https://doi.org/10.1002/asi.21173
3. Alwedyan, J., Hadi, W.M., Salam, M., Mansour, H.Y.: Categorize Arabic data sets using multi-class classification based on association rule approach. In Proceedings of the 2011 International Conference on Intelligent Semantic Web-Services and Applica-tions, pp. 1–8. Association for Computing Machinery, New York (2011). https://doi.org/10.1145/1980822.1980840
4. Hakak, S., Kamsin, A., Zada Khan, W., Zakari, A., Imran, M., bin Ahmad, K., Amin Gilkar, G.: Digital Hadith authentication: Recent advances, open challenges, and fu-ture directions. Transactions on Emerging Telecommunications Technologies (2020). https://doi.org/10.1002/ETT.3977
5. Najeeb, M.M.A.: Towards a deep learning-based approach for Hadith classification. European Journal of Engineering and Technology Res. **6**(3), 9–15 (2021). https://doi.org/10.24018/ejers.2021.6.3.2378
6. Ghazizadeh, M., Zahedi, M.H., Kahani, M., Bidgoli, B.M.: Fuzzy expert system in determining Hadith validity. In Sobh, T. (ed.) Advances in Computer and

Information Sciences and Engineering, pp. 354–359. Springer, Dordrecht (2008). https://link.springer.com/chapter/10.1007/978-1-4020-8741-7_64
7. Bilal, K., Mohsin, S.: Muhadith: A cloud based distributed expert system for classification of Ahadith. In 2012 10th International Conference on Frontiers of Information Technology, pp. 73–78, IEEE, New York (2012). https://doi.org/10.1109/FIT.2012.22
8. Aldhaln, K., Zeki, A., Zeki, A.: Knowledge extraction in Hadith using data mining technique. Int. J. Inf. Technol. Comput. Sci. **2**, 13–21 (2012).
9. Najiyah, I., Susanti, S., Riana, D., Wahyudi, M.: Hadith degree classification for Shahih Hadith identification web based. In 2017 5th International Conference on Cyber and IT Service Management, pp. 1–6, IEEE, New York (2017). https://doi.org/10.1109/citsm.2017.8089304
10. Brown, J.A.C.: Hadith: Muhammad's legacy in the medieval and modern world. One-world Publications, London (2017).
11. Kilgarriff, A., Grefenstette, G.: Introduction to the special issue on the web as corpus. Comput. Linguist. **29**(3), 333–347 (2003). https://doi.org/10.1162/089120103322711569
12. Cohen, J.: A coefficient of agreement for nominal scales. Educ. Psychol. Meas. **20**(1), 37–46 (1960). https://doi.org/10.1177/001316446002000104
13. Landis, J.R., Koch, G.G.: The measurement of observer agreement for categorical data. Biometrics. **33**(1), 159–174 (1977). https://doi.org/10.2307/2529310
14. Tarmom, T., Atwell, E., Alsalka, M.: Non-authentic Hadith corpus: Design and methodology. International Conference on Islamic Applications in Computer Science and Technologies. (2019). http://www.sign-ific-ance.co.uk/index.php/IJASAT/article/view/2272
15. Altammami, S., Atwell, E., Alsalka M.A.: The Arabic–English parallel corpus of authentic Hadith. International Conference on Islamic Applications in Computer Science and Technologies. (2019). http://www.sign-ific-ance.co.uk/index.php/IJASAT/article/view/2199
16. Zhang, J., Li, Y., Tian, J., Li, T.: LSTM–CNN hybrid model for text classification. In 2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), pp. 1675–1680, IEEE, New York (2018). https://doi.org/10.1109/IAEAC.2018.8577620
17. Li, X., Ning, H.: Chinese text classification based on hybrid model of CNN and LSTM. In Proceedings of the 3rd International Conference on Data Science and Information Technology, pp. 129–134, Association for Computing Machinery, New York (2020). https://doi.org/10.1145/3414274.3414493
18. Teahan, W.J.: A compression-based toolkit for modelling and processing natural language text. Information **9**(12), 294 (2018). https://doi.org/10.3390/info9120294
19. Cleary, J., Witten, I.: Data Compression Using Adaptive Coding and Partial String Matching. In IEEE Transactions on Communications, vol. 32, no. 4, pp. 396-402, April (1984). https://doi.org/10.1109/TCOM.1984.1096090
20. Daniel, J., Martin, J.H.: Speech and Languages Processing. Prentice-Hall, New Delhi (2000)