Computers, Materials & Continua
DOI:10.32604/cmc.2021.015865
Article

Tech Science Press

Entropy-Based Watermarking Approach for Sensitive Tamper Detection of Arabic Text

Fahd N. Al-Wesabi[1,2,*]

[1]Department of Computer Science, King Khalid University, Muhayel Aseer, Kingdom of Saudi Arabia
[2]Faculty of Computer and IT, Sana'a University, Sana'a, Yemen
[*]Corresponding Author: Fahd N. Al-Wesabi. Email: falwesabi@kku.edu.sa
Received: 11 December 2020; Accepted: 11 January 2021

Abstract: The digital text media is the most common media transferred via the internet for various purposes and is very sensitive to transfer online with the possibility to be tampered illegally by the tampering attacks. Therefore, improving the security and authenticity of the text when it is transferred via the internet has become one of the most difficult challenges that researchers face today. Arabic text is more sensitive than other languages due to Harakat's existence in Arabic diacritics such as Kasra, and Damma in which making basic changes such as modifying diacritic arrangements can lead to change the text meaning. In this paper, an intelligent hybrid solution is proposed with highly sensitive detection for any tampering on Arabic text exchanged via the internet. Natural language processing, entropy, and watermarking techniques have been integrated into this method to improve the security and reliability of Arabic text without limitations in text nature or size, and type or volumes of tampering attack. The proposed scheme is implemented, simulated, and validated using four standard Arabic datasets of varying lengths under multiple random locations of insertion, reorder, and deletion attacks. The experimental and simulation results prove the accuracy of tampering detection of the proposed scheme against all kinds of tampering attacks. Comparison results show that the proposed approach outperforms all of the other baseline approaches in terms of tampering detection accuracy.

Keywords: Entropy; text features; tamper detection; arabic text; NLP

1 Introduction

Nowadays, most digital media exchanged via the internet is in text form. The security challenges of text media have increased especially in terms of tampering by illegal attacks during the transmission process over the internet [1]. The reliability and security of text media have been improved regarding several information security techniques for preserving authentication of information transmission [2]. Several solutions and algorithms have been proposed for various purposes of digital media security such as copyright protection, verification of integrity, identification of owners, and access control. Most of these solutions are limited and applicable solely for images, audio, and video media in which hiding secure information within pixels of images, waves of

audio, and frames of video to validate their authenticity later after transmission is easy, however, there is lack of such applicable solutions for text media [3–5].

A technique of digital watermarking (DWM) is one of the most used techniques for text media security, which can be embedded inside a digital text to provide security and reserve the watermarked text to have an authentic text without change [6].

Three core issues must be resolved in the text watermarking area: accuracy, security, and robustness [7]. Proposing high reliable approaches and strategies for sensitive text documents, especially in Arabic and English languages, is the most common challenge in this area [8]. Therefore, authentication of content, verification of honesty, and detection of tampering of sensitive text are major issues in different systems that require critical solutions [9].

The common critical examples of such critical sensitive digital contents are the Arabic Holy Qur'an, eChecks, online tests, and marks. Different Arabic alphabet characteristics such as diacritic lengthened letters and extra symbols of Arabic make it simple to modify the key meaning of the text material by making basic changes such as modifying diacritic arrangements [10,11]. The most popular soft computation and natural language processing (NLP) techniques that support the analysis of the text are hidden Markov model (HMM) and Entropy analysis.

In this paper, a hybrid method has been proposed for improving the detection accuracy of illegal tampering of Arabic text exchanged via the internet. The proposed method uses NLP and entropy analysis to analyze the given Arabic text and extract its interrelationship features. In addition, it utilizes the extracted features of the given Arabic text as a watermark key by embedding them logically within the original text and checks the authenticity of the given Arabic text after the transmission process. As a result, when any illegal tampering attacks are detected, one can then truly say whether the text was, indeed, tampered. The main contribution of the proposed method is to meet the high accuracy of identification of sensitive tampering attacks that can occur on the Arabic text transmitted through the internet.

The rest of the paper has four more sections. Section 2 presents a related work. Section 3 provides the proposed method. Section 4 describes the implementation, simulation, comparison, and results discussion and Section 5 offers conclusions.

2 Related Work

According to the literature reviewed in this paper of text watermarking area, the existing methods and solutions are classified into linguistic, structural, and zero-watermark methods [1,3,12]

In [12], a zero-watermarking technique has been proposed to tamper detection in the plain text based on the quality of the given context which can later be obtained using the extraction algorithm to define the tampering position in the text document. A zero-watermarking scheme has been presented in [13] for authentication purpose using the properties of the Arabic alphabets as series of verses without changing the original file. The proposed scheme applied to Holy Quran authentication in which the watermark key is produced by checking the surah name, number and verse numbers and then matching them with the one stored at the certification authority. A new tamper locating technique has been proposed in [14] to authenticate the contents of DOCX documents. The key mechanism of the proposed technique is storing the authentication information in the key file with .xml extension based on the segmentation of the displayed letters.

To enhance the imperceptibility and capacity of the Arabic text, a method of text watermarking suggested based on the accessible words [15]. In this method, any word-space is used to mask the Boolean bit 0 or 1 that physically modifies the original text. A new text-based steganography scheme was proposed to hide secure data in the Arabic documents [16]. The mechanism of this scheme considers Harakat's and features exist in Arabic diacritics such as Kasra, Fatha, and Damma as well as reverses Fatha to cover the message.

A Kashida-marks invisible method of watermarking [17], based on the features of frequent recurrence of document security and authentication characters, was proposed. The method is based on a predetermined watermark key with a Kashida placed for a bit 1 and a bit omitted.

The method of steganography of the text was proposed to use Kashida extensions depends on the characters 'moon' and 'sun' to write digital contents of the Arabic language [18]. Also, Kashida characters are seen alongside with characters from Arabic to decide which hidden secret bits are kept by specific characters. In this form, four instances are included in the Kashida characters: moon characters representing '00'; sun characters representing '01'; sun characters representing '10'; and moon characters representing '11'. In ref. [19], a Kashida-based algorithm has been presented to enhance the capacity and security issues based on the insertion of variable numbers of Kashida per word. The proposed algorithm was designed to hiding the secret information within the texts using the Arabic script features based on the extension of the Kashida.

A Unicode extended characters scheme has been presented in [20] to secure the textual information from illegal attacks. The mechanism of the proposed scheme depends on three main steps, the development, incorporation, and extraction of the watermark. The added watermark key is focused on the development of predefined coding tables. However, the scrambling strategies are often used to generate or remove the watermark key in safe mode. The substitution-based attack method has been proposed in [21] based on preserving the position of words in the text document. This method depends on manipulating word transitions in the text document.

Zero-based watermarking models have been proposed in [22,23] for authenticating information and detecting the malicious tampering of the Arabic text media. The proposed models use the Markov model to analyze the Arabic text and extract its characteristics to utilize them as watermark data. The findings of the experiment of the proposed models provide good efficiency and sensitivity to all kinds of tamper attacks that can occur randomly in any position of the Arabic text. Chinese text-based watermarking methods have been proposed in [24,25] for authentication of Chinese text documents based on the combination of the properties of sentences. The proposed methods are presented as follows: a text of the Chinese language is split into a group of sentences, and for each word, the code of semantic has been obtained. The distribution of semantic codes influences sentence entropy.

In reference [26], zero-watermarking technique has been presented to resolve the security issues of English text-documents such as verification of content and copyright protection. A zero-watermarking methods have been suggested based on Markov-model authentication of the content of English text [27,28]. In these methods, to extract the safe watermark information, the probability characteristics of the English text are involved and stored to confirm the validity of the attacked text-document. These methods provide good security against popular text attacks. For the defense of English text copyright, based on the present rate of ASCII non-vowel letters and terms, the conventional watermark technique [29] has been suggested.

All those approaches review in the literature of this paper make use of traditional techniques and lack addressing solutions for Arabic text in terms of tampering detection issues due to the complexity, nature, and structure of the Arabic text. However, the proposed method is an efficient hybrid model which makes use of both mentioned mechanisms and addresses the problem of content authentication and tampering detection even on very low tampering.

3 Proposed Method

An intelligent method has been proposed in this paper named (Intelligent Hybrid Scheme for Tamper Detection of Arabic Text) and abbreviated as IHSTDAT. NLP, Entropy, and zero watermarking techniques have been integrated with IHSTDAT to improve the detection accuracy of tampering attacks. The proposed method involved four core steps starting with a pre-processing step. The second step includes the Arabic text feature extraction and watermark generation. The third step involves the attack process and watermark extraction. The final step is the matching process. The details of these steps are explained below.

3.1 Pre-Processing Step

Spaces and newlines in the Arabic text can directly affect the meaning of the Arabic text and accuracy results. This step is responsible of checking the existing spaces and newlines within the given text and creating or removing them as required for the Arabic text. This step should be performed for both the original and attacked Arabic text. The output of this step is the pre-processed original or attacked Arabic text which will be used as an input for the next step (Arabic text analysis and feature extraction) as illustrated in Fig. 1 below.

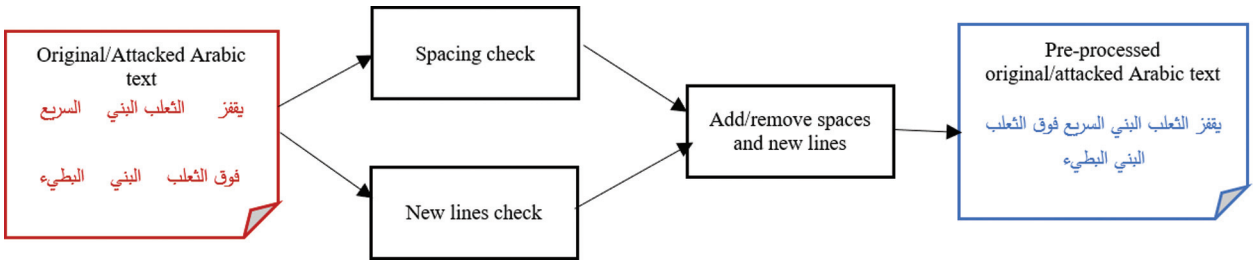

Figure 1: Pre-processing step of IHSTDAT

3.2 Feature Extraction and Watermark Generation Step

The proposed method is based on text analysis and feature extraction of the given text. These extracted features represent the interrelationships between the contexts of the given text which will be used as a watermark key. Watermark embedded has taken place logically in this method without any need to change the original text. This step can be divided into three sub-processes: text analysis, feature extraction, and watermark generation and embedding. The details of these sub-activities are explained below by using the following Arabic text sample.

"يقفز الثعلب البني السريع فوق الثعلب البني البطيء للوصول إلى الثعلب البني الميت"

3.2.1 Arabic Text Analysis and Feature Extraction

This sub-process represents a core activity of the proposed method. Text analysis will be performed using the Markov model to analyze the context of the given Arabic text. The pre-processed original Arabic text is required as an input for this subprocess. Features of the given Arabic text will be extracted as a result of this process. The extracted features refer to the interrelationship between the given Arabic text. Three processes should be performed in this step: building Markov chain matrix, representing the given Arabic contexts as unique states and transitions, and finally finding the interrelationship between the given Arabic contexts. The available states and their transitions in the Arabic text sample are illustrated in Fig. 2 below.

```
                                                    }
                ("يقفز")    : [ ' الثعلب'] ،
("الثعلب")    : [ ' البني' ، ' البني' ، ' البني' ]،
("البني")     : [ ' السريع ' ، ' البطيء' ، ' الميت ' ]،
("السريع")    : [' فوق' ]،
("البطيء")    : [ ' للوصول ' ] ،
("للوصول") : [' إلى ' ] ،
                                                    {
```

Figure 2: The available states and their transitions using IHSTDAT

The author assumes that "بلعثال" is a current state, and its next transitions are "ينبال," "ينبال," and "ينبال." It is observed that ينبال transition occurs three times.

Building the Markov chain matrix and using it for representing the states and transitions of the given Arabic contexts and initializing zero value for all available transitions are shown in Fig. 3.

States	Transitions								
	يقفز	الثعلب	البني	السريع	فوق	البطيء	للوصول	إلى	الميت
يقفز	0	0	0	0	0	0	0	0	0
الثعلب	0	0	0	0	0	0	0	0	0
البني	0	0	0	0	0	0	0	0	0
السريع	0	0	0	0	0	0	0	0	0
فوق	0	0	0	0	0	0	0	0	0
البطيء	0	0	0	0	0	0	0	0	0
للوصول	0	0	0	0	0	0	0	0	0
إلى	0	0	0	0	0	0	0	0	0

Figure 3: Representation of the available states and their transitions using IHSTDAT

The text analysis and feature extraction of the given Arabic text sample are performed by the proposed method IHSTDAT to obtain the interrelationship of the given Arabic contents by computing the number of appearances of the potential conversions for every present state of a

single unique Arabic word as calculated by Eq. (1) and illustrated in Fig. 4.

$$fe[ps][ns] = \sum_{i,j=0}^{n-1} ts[i][j] \tag{1}$$

Which n is a total number of available states and ts refers to the total of transitions.

States	Transitions									
	يقفز	الثعلب	البني	السريع	فوق	البطيء	للوصول	إلى	الميت	Entropy value
يقفز	0	1	0	0	0	0	0	0	0	
الثعلب	0	0	3	0	0	0	0	0	0	
البني	0	0	0	1	0	1	0	0	1	
السريع	0	0	0	0	1	0	0	0	0	
فوق	0	1	0	0	0	0	0	0	0	
البطيء	0	0	0	0	0	0	1	0	0	
للوصول	0	0	0	0	0	0	0	1	0	
إلى	0	1	0	0	0	0	0	0	0	
Watermark key										

Figure 4: Text analysis and feature extraction of the provided Arabic-text using IHSTDAT

3.2.2 Watermark Generation and Embedding

The results of the text analysis and feature extraction process are utilized as inputs to this process in which entropy analysis will be applied to them for all transitions of each state by Eq. (2). However, the summation of all entropy values represents the watermark key which is calculated by Eq. (3) and illustrated in Fig. 5.

$$Entropy = -\sum_{i=1}^{n} s_i log_2 (s_i) \tag{2}$$

where s_i is a transition(s) of each state.

$$ERWM = -\sum_{i,j=1}^{n} Entropy[i][j] \tag{3}$$

In this method, the extracted features of the given Arabic-text and the generated watermark key are embedded logically with no need to change the original text by identifying all non-zero values in the Markov chain matrix.

3.3 Attack Process and Watermark Extraction Step

The core input of this step is the pre-processed attacked Arabic text. This process is a reverse process of the embedding process which is responsible for extracting the Arabic text features that are embedded logically within both the original and the attacked given Arabic text. The outputs of this process will be used as inputs for the matching process in the next step. The algorithm of attack and watermark extraction step is illustrated below in Algorithm 1.

Algorithm 1: The algorithm of attack and watermark extraction step using IHSTDAT

PROCEDURE extraction (OAT_P)
1. Input: the attacked Arabic text (OAT_P)
2. Output: the attacked watermark (WMK_A)
3. BEGIN
4. Watermark
5. **for** i = 1 to entropy_mm.length – 1
6. **for** j = 1 to entropy_mm.length
7. **If** entropy_mm[i][j] != 0
8. WMK_A += entropy_mm[i][j]
9. **return** WMK_A

States	Transitions									
	يقفز	الثعلب	البني	السريع	فوق	البطيء	للوصول	إلى	الميت	Entropy value
يقفز	0	1	0	0	2	0	0	0	0	0.918
الثعلب	0	0	3	0	0	0	1	0	0	0.811
البني	0	0	0	1	0	1	0	0	1	1.585
السريع	0	0	0	0	1	0	0	0	0	0.269
فوق	0	1	0	0	0	0	0	0	0	0.269
البطيء	0	0	1	0	1	0	1	0	1	2
للوصول	0	0	0	0	0	0	0	1	0	0.269
إلى	0	1	0	0	0	0	0	0	0	0.269
Watermark key										5.314

Figure 5: Entropy-analysis and generating watermark process of IHSTDAT

3.4 Matching Step

The extracted features of both the original and the attacked of the given Arabic text should be provided as inputs to run this algorithm. However, the status of the given Arabic-text is a core output of this step which can be authentic or tampered. The matching process is performed by two levels which are:

- *1st level*: matching the whole original and attacked watermark keys. If these two watermark keys are similar in values and appearance, then the matching process will be stopped and the authentic Arabic text will be notified.
- *2nd level*: matching based on the state level: this matching will be performed for each entropy value of every state's transition which is calculated by Eq. (4).

$$ERM_s(i,j) = \left| \frac{ER_O[i][j] - (ER_O[i][j] - ER_A[i][j])}{ER_O[i][j]} \right| \tag{4}$$

where ERM_S is the matching rate at the entropy of state level, ER_O is the original entropy rate, and ER_A is the attacked entropy rate.

The accuracy of the tampering detection is computed by Eq. (5) below.

$$ERaccuracy = \frac{\sum_{i=1}^{n}(ERMs)}{total\ number\ of\ transitions} * 100 \quad (5)$$

The matching step is illustrated in Fig. 6.

In addition, Algorithm 2 shows the whole pseudocode scheme of the proposed method IHSTDAT.

Algorithm 2: The general algorithm of the proposed IHSTDAT method

PROCEDURE IHSTDAT

```
Input: Original Arabic text (OAT)
Output: Tampering detection accuracy (ERaccuracy), authentication notification
BEGIN
// perform pre-processing process
pre-processing (OAT)
// perform Arabic text analysis and feature extraction process
TAFE(OATP)
// perform Entropy analysis and watermark generation process
entropy_WM_gen (TAFE(OATP))
// perform extraction process for both OATO and OATP
extraction (OATO , OATP)
// perform matching process
 IF WMKO = WMKP
      Print "An authentic Arabic text"
      ERM = 100
 ELSE
      Print "Tampered Arabic text"
// compute entropy matching rate on transition level
      for i = 1 to entropy_mm.length - 1,
          for j = 1 to entropy_mm.length
              IF ERO[i][j] != 0
                  patternCount  +=1
                  ERMs(i,j) = |(ERO[i][j] - (ERO[i][j] - ERA[i][j])) / ERO[i][j]|
                  transERMtotal += ERMS
              ELSE
                    IF ERA[i][j] != 0
                        patternCount += ERA[i][j]
// compute accuracy of entropy matching rate on a whole given text
ERaccuracy = (∑_{i=1}^{n}(ERMs)) / (Total number of transitions) * 100
return ERaccuracy
```

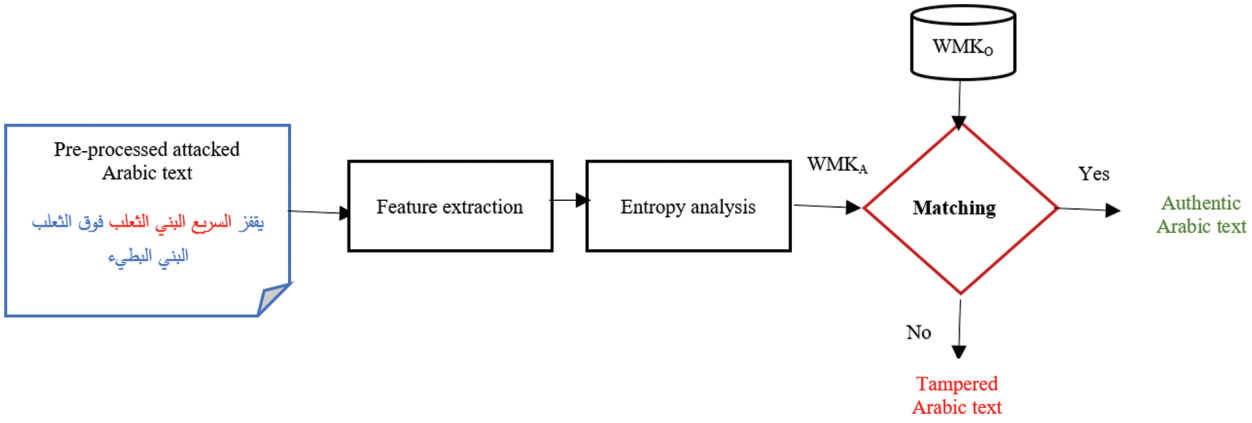

Figure 6: Matching step using IHSTDAT

4 Implementation, Results Discussion, and Comparison

The implementation of the proposed method IHSTDAT has been performed by self-developed software via the use of PHP vs code. Different simulation scenarios have been carried out using standard Arabic datasets of various sizes categorized as small, mid, and large datasets under predefined random attacks with their volumes categorized as low, mid, and high volumes.

4.1 Simulation Results of IHSTDAT

To evaluate the accuracy of tampering detection of IHSTDAT, scenarios of many studies are performed as shown in Tab. 1, for all forms of attacks and their volumes.

Table 1: Detection accuracy evaluation of IHSTDAT

Dataset (%)	Attacks		
	Insertion	Deletion	Reorder
5	93.34	90.97	87.97
10	89.93	83.67	78.10
20	81.42	63.66	65.99
50	65.09	35.87	45.56

From Tab. 1 above and Fig. 7 below, it seems that the IHSTDAT approach gives sensitive results of detection of tampering in all attacks that the structure, semantics, and syntax of the content of Arabic text may have been carried out. As a comparison of detection accuracy based on attack types, the results show that the most sensitive tampering detection in all attack volume scenarios is the insertion attack.

4.2 Comparison Results

This subdivision provides a comparison of the proposed IHSTDAT methods with other baseline methods named HNLPZWA presented in [22] and ZWAFWMM presented in [23]. The comparison analyzes their results in conjunction with key affected variables (i.e., scale of databases, forms of attacks, and volumes).

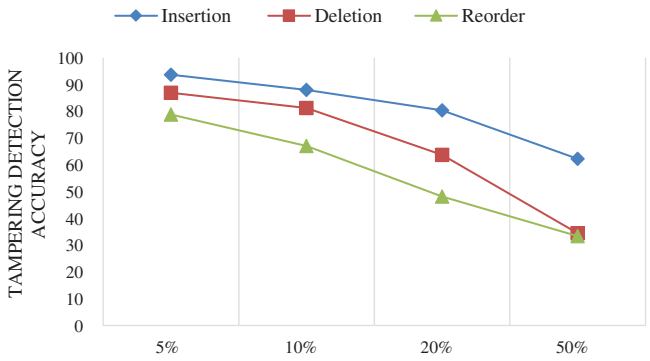

Figure 7: Accuracy evaluation of IHSTDAT using all attack's rates

4.2.1 Comparison Results Under Low Attack Rate

Tab. 2 shows a comparison of the low attack rate effect on tampering detection accuracy of the proposed IHSTDAT method and other baseline methods HNLPZWA and ZWAFWMM.

Table 2: Detection accuracy comparison under low attack rate

Approach	Attack type		
	Insertion	Deletion	Reorder
IHSTDAT	93.57	86.85	78.79
ZWAFWMM	92.28	86.51	72.00
HNLPZWA	91.37	85.58	69.33

As a comparison of the general performance of all compared methods based on attack type, the results shown in Tab. 2 and Fig. 8 show that the insertion attack is the most sensitive for tampering and provides better accuracy than deletion and reorder attacks in all comparison scenarios. As a comparison of detection accuracy based on the compared methods, it seems that the proposed IHSTDAT method outperforms other baseline methods in all scenarios. This means the proposed IHSTDAT method is very sensitive to detect any tampering on Arabic text whenever the tampering volume is very low.

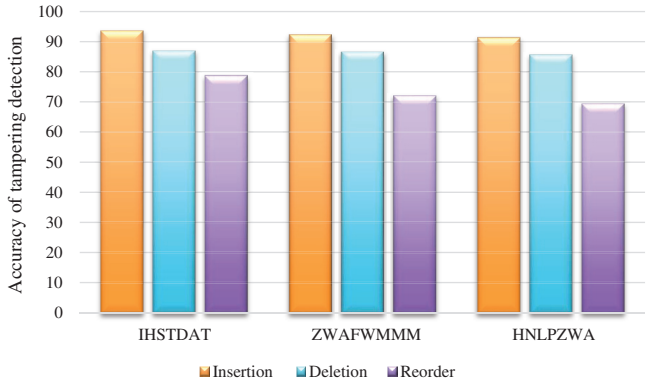

Figure 8: Detection accuracy comparison under low attack rate

4.2.2 Comparison Results Under Mid Attack Rate

Tab. 3 shows a comparison of the mid-attack volume effect on tampering detection accuracy of the proposed IHSTDAT method and other baseline methods HNLPZWA and ZWAFWMM.

Table 3: Detection accuracy comparison under mid-attack rate

Approach	Attack type		
	Insertion	Deletion	Reorder
IHSTDAT	80.27	63.66	48.18
ZWAFWMM	79.67	63.22	35.28
HNLPZWA	78.60	62.08	32.46

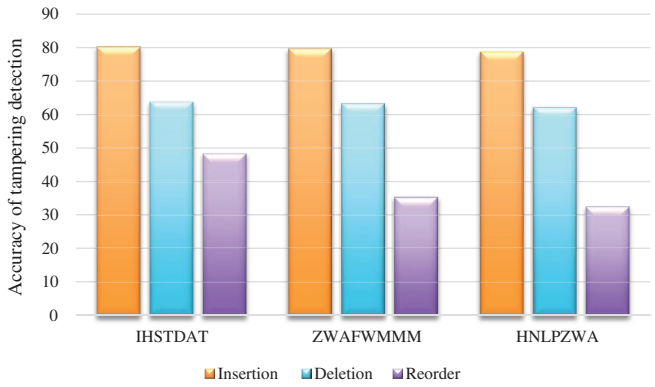

Figure 9: Detection accuracy comparison under mid-attack rate

According to the attack type factor, the results shown in Tab. 3 and Fig. 9 show that the insertion attack gives the best detection accuracy in all comparison scenarios. As a comparison of detection accuracy based on the compared methods, it seems the proposed IHSTDAT method outperforms other baseline methods in all scenarios. This means the proposed IHSTDAT method is very sensitive to detect any tampering on Arabic text whenever the tampering volume is very low or mid.

4.2.3 Comparison Results Under High Attack Rate

Tab. 4 shows a comparison of the high attack rate effect on tampering detection accuracy of the proposed IHSTDAT method and other baseline methods HNLPZWA and ZWAFWMM.

Tab. 4 and Fig. 10 demonstrate how the accuracy of tampering detection is affected under the high volume of attacks against all compared methods. According to the attack type factor, the results show that the insertion attack gives the best detection accuracy as the same cases of low and mid volumes of attacks. As a comparison of detection accuracy based on the compared methods, results show that the proposed IHSTDAT method outperforms other baseline methods in all scenarios. This means the proposed IHSTDAT method is very sensitive to detect any tampering on Arabic text whenever tampering volume is very low, mid or high.

Table 4: Detection accuracy comparison under high attack rate

Approach	Attack type		
	Insertion	Deletion	Reorder
IHSTDAT	62.21	34.42	33.40
ZWAFWMM	60.89	36.12	15.66
HNLPZWA	60.33	34.39	13.66

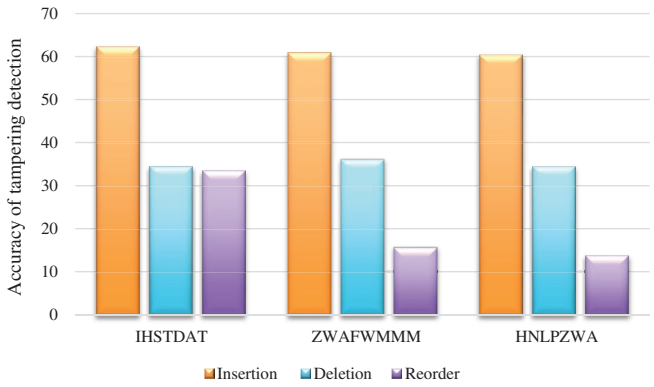

Figure 10: Detection accuracy comparison under high attack rate

5 Conclusion

In this paper, an intelligent method has been proposed to enhance the accuracy of tampering detection of Arabic text based on hybrid techniques (entropy, NLP, and watermarking techniques). The contribution of the proposed method is to meet high tampering detection accuracy of sensitive Arabic text exchanged via the internet without limitations in contents, nature, structure, or size of Arabic text. The types and volumes of tampering attacks are also addressed by the proposed method which has no limitations in the type of tampering attacks (insertion, deletion, or reorder attacks with very low, low, mid, or high volumes). The proposed method IHSTDAT has been implemented using self-developed software. However, the performance and detection accuracy of IHSTDAT have been simulated and evaluated under various scenarios of simulation and various regular Arabic datasets by different amounts of attacks. The baseline techniques HNLPZWA and ZWAFWMM were compared to the proposed IHSTDAT method. The findings reveal that IHSTDAT beats HNLPZWA and ZWAFWMM in terms of general performance and tampering detection accuracy. Furthermore, the findings illustrate that IHSTDAT refers to all Arabic literature, numbers, spaces, and special characters. For future research, the enhancement of detection accuracy and watermark fragility for all kinds of attacks should be considered.

Funding Statement: The author extends his appreciation to the Deanship of Scientific Research at King Khalid University for funding this work under Grant Number (R. G. P. 2/55/40/2019), Received by Fahd N. Al-Wesabi. www.kku.edu.sa.

Conflicts of Interest: The author declares that he has no conflicts of interest to report regarding the present study.

References

[1] F. N. Al-Wesabi, "A smart English text zero-watermarking approach based on third-level order and word mechanism of Markov model," *Computers, Materials & Continua*, vol. 65, no. 2, pp. 1137–1156, 2020.

[2] M. Fujimura, K. Imamura and H. Kuroda, "Application of saliency map to restraint scheme of attack to digital watermark using seam carving," in *IEEE Int. Conf. on Consumer Electronics-Taiwan*, Taipei, Taiwan. IEEE, pp. 347–348, 2017.

[3] F. N. Al-Wesabi, "A hybrid intelligent approach for content authentication and tampering detection of Arabic text transmitted via internet," *Computers, Materials & Continua*, vol. 66, no. 1, pp. 195–2011, 2021.

[4] J. Abraham and V. Paul, "An imperceptible spatial domain color image watermarking scheme," *Journal of King Saud University-Computer and Information Sciences*, vol. 31, no. 1, pp. 125–133, 2019.

[5] J. Mayer, P. V. Borges and S. J. Steven, *Fundamentals and Applications of Hardcopy Communication*. Brazil: Springer, pp. 1–5, 2018.

[6] R. A. Alotaibi and L. A. Elrefaei, "Improved capacity Arabic text watermarking methods based on open word space," *Journal of King Saud University-Computer and Information Sciences*, vol. 30, no. 2, pp. 236–248, 2018.

[7] S. Hakak, A. Kamsin, O. Tayan, M. Y. Idris, A. Gani *et al.,* "Preserving content integrity of digital holy Quran: Survey and open challenges," *IEEE Access*, vol. 5, no. 10, pp. 7305–7325, 2017.

[8] R. Ahmed and L. Elrefaei, "Arabic text watermarking: A review," *International Journal of Artificial Intelligence & Applications*, vol. 6, no. 4, pp. 1–16, 2015.

[9] A. A. Alwan, M. S. Abdulah and N. N. Sjarif, "A survey on combined various data hiding techniques," *Open International Journal of Informatics*, vol. 7, no. 2, pp. 31–44, 2019.

[10] S. Hakak, A. Kamsin, O. Tayan, M. Yamani and G. Gilkar, "Approaches for preserving content integrity of sensitive online Arabic content: A survey and research challenges," *ELSEVIER Information Processing and Management*, vol. 56, no. 2, pp. 367–380, 2019.

[11] K. Hameed, A. Khan, M. Ahmed and A. G. Reddy, "Towards a formally verified zero watermarking scheme for data integrity in the internet of things based-wireless sensor networks," *ELSEVIER Future Generation Computer Systems*, vol. 167, pp. 1–16, 2018.

[12] Z. Jalil, A. M. Mirza and M. Sabir, "Content based zero-watermarking algorithm for authentication of text documents," *arXiv preprint arXiv*, vol. 7, no. 2, pp. 212–217, 2010.

[13] Y. M. Alginahi, M. N. Kabir and O. Tayan, "An enhanced Kashida-based watermarking approach for Arabic text-documents," in *Int. Conf. on Electronics, Computer and Computation*, Ankara, Turkey, pp. 301–304, 2013.

[14] H. Hai, X. D. Qing and Q. Ke, "A watermarking-based authentication and image restoration in multimedia sensor networks," *International Journal of High-Performance Computing and Networking*, vol. 12, no. 1, pp. 65–73, 2018.

[15] R. Alotaibi and L. Elrefaei, "Improved capacity Arabic text watermarking methods based on open word space," *Journal of King Saud University–Computer and Information Sciences*, vol. 30, no. 2, pp. 236–248, 2018.

[16] M. Memon and A. Shah, "A novel text steganography technique to Arabic language using reverse Fat5Th5Ta," *Pakistan Journal of Engineering, Technology and Sciences*, vol. 1, no. 2, pp. 106–113, 2015.

[17] Y. Alginahi, M. Kabir and O. Tayan, "An enhanced Kashida-based watermarking approach for increased protection in Arabic text-documents based on frequency recurrence of characters," *International Journal of Computer and Electrical Engineering*, vol. 6, no. 5, pp. 381–392, 2014.

[18] A. Shaker, F. Ridzuan and S. Pitchay, "Text Steganography using extensions kashida based on the moon and sun letters concept," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 8, pp. 286–290, 2017.

[19] M. H. Kamarudin, C. Maple and T. Watson, "Hybrid feature selection technique for intrusion detection system," *International Journal of High-Performance Computing and Networking*, vol. 13, no. 2, pp. 232–240, 2019.

[20] N. Al-maweri, W. Adnan, A. Rahman, S. Khairulmizam and S. Syed, "Robust digital text watermarking algorithm based on unicode extended characters," *Indian Journal of Science and Technology*, vol. 9, no. 48, pp. 1–14, 2016.

[21] M. Bashardoost, M. Rahim, T. Saba and A. Rehman, "Replacement attack: A new zero text watermarking attack," *3D Research*, vol. 8, no. 1, pp. 11, 2017.

[22] F. N. Al-Wesabi, "Proposing high-smart approach for content authentication and tampering detection of Arabic text transmitted via Internet," *IEICE transactions in Information Systems*, vol. E103, no. 10, pp. 2104–2112, 2020.

[23] F. N. Al-Wesabi, K. Mahmood and N. Nemri, "A zero watermarking approach for content authentication and tampering detection of Arabic text based on fourth level order and word mechanism of Markov model," *Journal of Information Security and Applications*, vol. 52, no. 10, pp. 1–15, 2020.

[24] Y. Liu, Y. Zhu and G. Xin, "A zero-watermarking algorithm based on merging features of sentences for Chinese text," *Journal of the Chinese Institute of Engineers*, vol. 38, no. 3, pp. 391–398, 2015.

[25] P. Zhu, G. Xiang, W. Song, A. Li, Y. Zhang *et al.,* "A text zero watermarking algorithm based on Chinese phonetic alphabets," *Wuhan University Journal of Natural Sciences*, vol. 21, no. 4, pp. 277–282, 2016.

[26] O. Tayan, Y. Alginahi and M. Kabir, "An adaptive zero-watermarking approach for text documents protection," *International Journal of Image Processing Techniques*, vol. 1, no. 1, pp. 33–36, 2014.

[27] M. Ghilan, F. Ba-Alwi and F. N. Al-Wesabi, "Combined Markov model and zero watermarking techniques to enhance content authentication of Arabic text documents," *International Journal of Computational Linguistics Research*, vol. 5, no. 1, pp. 26–42, 2014.

[28] F. N. Al-Wesabi, A. Alsakaf and K. U. Vasantrao, "A zero text watermarking algorithm based on the probabilistic patterns for content authentication of text documents," *International Journal of Computer Engineering & Technology*, vol. 4, no. 1, pp. 284–300, 2013.

[29] H. Ahmed and M. Khodher, "Comparison of eight proposed security methods using linguistic steganography text," *International Journal of Computing & Information Sciences*, vol. 12, no. 2, pp. 243–251, 2016.