Computers, Materials & Continua
DOI:10.32604/cmc.2021.018202
Article

Tech Science Press

An Optimized English Text Watermarking Method Based on Natural Language Processing Techniques

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: 01 March 2021; Accepted: 02 April 2021

Abstract: In this paper, the text analysis-based approach RTADZWA (Reliable Text Analysis and Digital Zero-Watermarking Approach) has been proposed for transferring and receiving authentic English text via the internet. Second level order of alphanumeric mechanism of hidden Markov model has been used in RTADZWA approach as a natural language processing to analyze the English text and extracts the features of the interrelationship between contexts of the text and utilizes the extracted features as watermark information and then validates it later with attacked English text to detect any tampering occurred on it. Text analysis and text zero-watermarking techniques have been integrated by RTADZWA approach to improving the performance, accuracy, capacity, and robustness issues of the previous literature proposed by the researchers. The RTADZWA approach embeds and detects the watermark logically without altering the original text document to embed a watermark. RTADZWA has been implemented using PHP with VS code IDE. The experimental and simulation results using standard datasets of varying lengths show that the proposed approach can obtain high robustness and better detection accuracy of tampering common random insertion, reorder, and deletion attacks, e.g., Comparison results with baseline approaches also show the advantages of the proposed approach.

Keywords: Text analysis; NLP; hidden markov model; zero-watermarking; content authentication; tampering detection

1 Introduction

For the research community, the reliability and security of exchanged text data through the internet is the most promising and challenging field. In communication technologies, authentication of content and automated text verification of honesty in different Languages and formats are of great significance. Numerous applications for instance; e-Banking and e-commerce. Render information transfer via the Internet the most difficult. In terms of content, structure, grammar, and semantics, much of the digital media transferred over the internet is in text form and is very susceptible to online transmission. During the transfer process, malicious attackers can temper such digital content and thus the changed count [1].

For information security, many algorithms and techniques are available such as the authentication of content, verification of integrity, detection of tampering, identification of owners, access control, and copyright protection.

To overcome these issues, steganography and automated methods of watermarking are commonly used. A technique of digital-Watermarking (DWM), which can be inserted into digital material through various details such as text, binary pictures, audio, and video [2,3]. A fine-grained text watermarking procedure is proposed based on replacing the white spaces and Latin symbols with homoglyph characters [4].

Several conventional methods and solutions for text watermarking were proposed [5,6] and categorized into different classifications such as linguistic, structure, image-based, and format-based images [7]. To insert the watermark information into the document, most of these solutions require certain upgrades or improvements to the original text in digital format material. Zero-watermarking without any alteration to the original digital material to embed the watermark information is a new technique with smart algorithms that can be used. Also, this technique can be used to generate data for a watermark in the contents of a given digital context [1,7–9].

Restricted research has centered on the appropriate solutions to verify the credibility of critical digital media online [10–12]. The verification of digital text and the identification of fraud in research earned great attention. In addition, text watermarking studies have research concentrated on copyright protection in the last decade. However, less interest and attention has been paid to integrity verification, identification of tampering, and authentication of content due to the existence of text content is natural language-dependent [13].

Proposing the most appropriate approaches and strategies for dissimilar formats and materials, especially in Arabic and English languages, is the most common challenge in this area [14,15]. Therefore, authentication of content, verification of honesty, and detection of tampering of sensitive text is a major issue in different systems that need critical solutions.

Some instances of such sensitive digital text content are Arabic interactive Holy Qur'an, online, eChecks, tests, and marks. Different Arabic alphabet characteristics such as diacritics 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 [16]. The most popular soft computation and natural language processing (NLP) technique that supported the analysis of the text is HMM.

The author suggests a reliable approach known as RTADZWA (Reliable Text Analysis and Digital Zero-Watermarking Approach) for transferring and receiving an authentic English text via the internet). The proposed approach is based on a second-order of alphanumeric mechanism based on the Markov model for content authentication and tampering detection of English text transmitted via the Internet. It consists of a model that operates in collaboration between zero watermarking and the Markov model as NLP techniques. In this approach, the second-order of alphanumeric mechanism has been used for text analysis in order to extract the interrelationships between the contents of the given English text and to generate a watermark key. The generated watermark will be embedded logically in the original English context without any modifications or effect on the size of the original text. Embedded watermark will be used later after the transmission of text via the Internet to detect any tampering occurring on the received English text and to determine if it is authentic or not.

The primary objective of the RTADZWA method is to achieve high accuracy of content authentication and sensitive detection of tampering attacks in English text, which has gained great importance and needs more security and protection via the Internet.

The remainder of the article is structured. In Section 2, the author explains the existing works done so far. In Section 3, the author discussed the suggested approach (RTADZWA). The simulation, implementation, are provided in Section 4, results discussion is provided in Section 5, and finally, the author concludes the article in Section 6.

2 Related Work

According to the processing domain of NLP and text watermarking, these existing methods and solutions of text watermarking reviewed in this paper classified into linguistical, structural, and zero-watermark methods [1,7,13].

2.1 Linguistical Methods

Natural language is the foundation of approaches to linguistic text watermarking. The mechanism of those methods embedding the watermark is based on changes applied to the semantic and syntactic essence of plain text [1].

To enhance the capability and imperceptibility of Arabic text, A method of text watermarking is suggested room dependent on the accessible words [17]. In this method, any word-space is used to mask the Boolean bit 0 or 1 that physically modifies the original text.

A text steganography technique was proposed to hide information in the Arabic language [18]. The step of this approach considers Harakat's existence 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 [19], 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 depend on the characters 'moon' and 'sun' to write digital contents of the Arabic language [20]. In addition, the method Kashida characters are seen alongside 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'.

2.2 Structural Methods

Structural methods are material dependent in which altering on the structure of the original text is performed to hide watermark data.

A text steganographic approach [21] based on multilingual Unicode characters has been suggested to cover details in English scripts for the use of the English Unicode alphabet in other languages. Thirteen letters of the English alphabet have been chosen for this approach. It is important to embed dual bits in a timeframe used ASCII code for embedding 00. However, multilingual ones were used by Unicode to embed between 01, and 10, as well as 11. The algorithm of Text Watermarking is used to secure textual contents from malicious attacks according to Unicode extended characters [22]. The algorithm requires three main steps, the development, incorporation, and extraction of watermarks. The addition of watermarks is focused on the development of predefined coding tables, while scrambling strategies are often used in generation and removal, the watermarking key is safe.

The substitution attack method focused on preserving the position of words in the text document has been proposed [23]. This method depends on manipulating word transitions in the text document. Authentication of Chinese text documents based on the combination of the properties of sentences, text-based watermarking approaches have been suggested [24,25]. The proposed method is presented as follows: firstly, a text of the Chinese language is split into a group of sentences, and for each word, the code of a semantic has been obtained. The distribution of semantic codes influences sentence entropy. The distribution of semantic codes influences sentence entropy.

2.3 Watermarking Methods

A zero-watermarking method has been proposed to preserve the privacy of a person who relies on the Hurst exponent and the nullity of the frames [26]. For watermark embedding, the two steps are determined to evaluate the unvoiced frames. The process of the proposed approach bases on integrating an individual's identity without notifying any distortion in the signals of medical expression.

A zero-watermarking method was proposed to resolve the security issues of text-documents of the English language, such as verification of content and copyright protection [27]. A zero-watermarking approach has been suggested based on the authentication Markov-model of the content of English text [28,29]. In this approach, to extract the safe watermark information, the probability characteristics of the text of English are involved and stored to confirm the validity of the attacked text-document. The approach provides security against popular text attacks with a watermark distortion rate if, for all known attacks, it is greater than one. For the defense of English text by copyright, based on the present rate of ASCII non-vowel letters and terms, the conventional watermark approach [30] has been suggested.

3 The Proposed Approach

This paper proposes a novel reliable approach by integrating NLP and text zero-watermark techniques in which there is no need to embed extra information such as watermark key, or even to perform any modifications to the original text. The second-order of alphanumeric mechanism of the Markov model has been used as NLP technique to analyze the contents of English text and extract the interrelationships features of these text contents.

The main contributions of our approach, RTADZWA can be summarized as follows:

- Unlike the previous work, in which the watermarking is performed by affecting text, content, and size, our approach RTADZWA embeds the watermarking logically without any effect on the text, content, and size.
- In our approach RTADZWA, watermarking does not need any external information because the watermark key is produced as a result of text analysis and extracting the relationship between the content itself and then making it as a watermark.

- Our approach RTADZWA is highly sensitive to any simple modification on the text and the meaning in the English text, which is known as the complex text. The three contributions mentioned above are found somehow only in images but not in the text. This is the vital point concerning the contribution of this paper.
- In addition, our approach RTADZWA can effectively determine the place of Tempering occurrence. This feature can be considered an advantage over the Hash function method.

3.1 Watermark Generation and Embedding Procedure

This subsection involves three sub-procedures which are pre-processing procedure, text analysis and watermark generation procedure, and watermark embedding procedure as illustrated in Fig. 1.

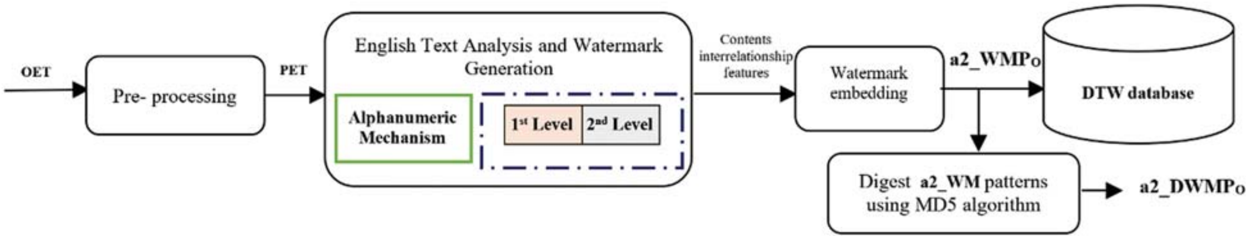

OET
Pre- processing
PET
English Text Analysis and Watermark Generation
Alphanumeric Mechanism
1st Level
2nd Level
Contents interrelationship features
Watermark embedding
a2_WMPo
DTW database
Digest a2_WM patterns using MD5 algorithm
a2_DWMPo

Figure 1: RTADZWA zero-watermark processes

3.1.1 Pre-processing Procedure

The pre-processing of the original English text is one of the key steps in both the watermark generation and extraction processes to convert letter cases from the capital to small letters, remove extra spaces and newlines, and it will directly influence the tampering detection accuracy and watermark robustness. The original English text (OET) is required as input for pre-processing process.

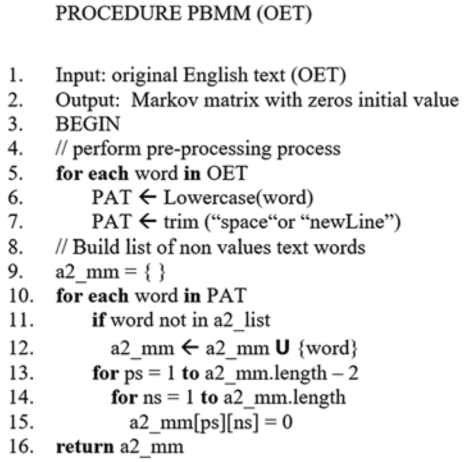
3.1.2 Text Analysis and Watermark Generation Procedure

This procedure includes two subprocesses that are building Markov matrix, text analysis, and watermark generation processes.

- *Building a Markov matrix* is the starting point of English text analysis and watermark generation process using the Markov model. A Markov matrix that represents the possible states and transitions available in a given text is constructed without reputations. In RTADZWA approach, each unique pair of alphanumeric within a given English text represents a present state, and each unique word a transition in the Markov matrix. During the building process of the Markov matrix, the proposed algorithm initializes all transition values by zero to use these cells later to keep track of the number of times that the i^{th} pair of alphanumeric is followed by the j^{th} alphanumeric within the given English text document.

The algorithm of the Markov matrix constructing is performed as shown in Algorithm 1 below.

Algorithm 1: Algorithm of building Markov matrix using RTADZWA



```
PROCEDURE PBMM (OET)

Input: original English text (OET)
Output:  Markov matrix with zeros initial value
BEGIN
// perform pre-processing process
for each word in OET
    PAT ← Lowercase(word)
    PAT ← trim (“space“or “newLine”)
// Build list of non values text words
a2_mm = { }
for each word in PAT
    if word not in a2_list
        a2_mm ← a2_mm U {word}
    for ps = 1 to a2_mm.length – 2
        for ns = 1 to a2_mm.length
            a2_mm[ps][ns] = 0
return a2_mm
```

where, OET: is an original English text, PET: is a pre-processed English text, a2_mm: states and transitions matrix with zeros values for all cells, ps: refers to the current state, ns: refers to next state.

Text analysis and watermark generation procedure: after the Markov matrix was constructed, natural language processing and text analysis process should be performed to found interrelationships between contexts of the given English text and generate watermark patterns. In this algorithm, the number of appearances of possible next states transitions for each current state of pair of alphanumeric will be calculated and constructed as transition probabilities by Eq. (1).

$$a2\_MM[ps][ns] = \sum_{i,j=1}^{n-2} Total\ Number\ of\ Transitions\,[i][j]\,, \tag{1}$$

where n: is the total number of states, and i: is i^{th} current state of pair of the alphanumeric.

This example of the English version demonstrates how this method was used to introduce the phase of transformation from the current state to the next state.

“The quick brown fox jumps over the brown fox who is slow jumps over the brown fox who is dead.”

When you use the second level of the secret Markov-model of alphanumeric approach, each pair of alphanumeric is a present state. Text processing is done as the text is read and the relationship meaning exchanged between the current and the next states is calculated. The accessible transitions from the above sample of the English text are shown in Fig. 2 below.

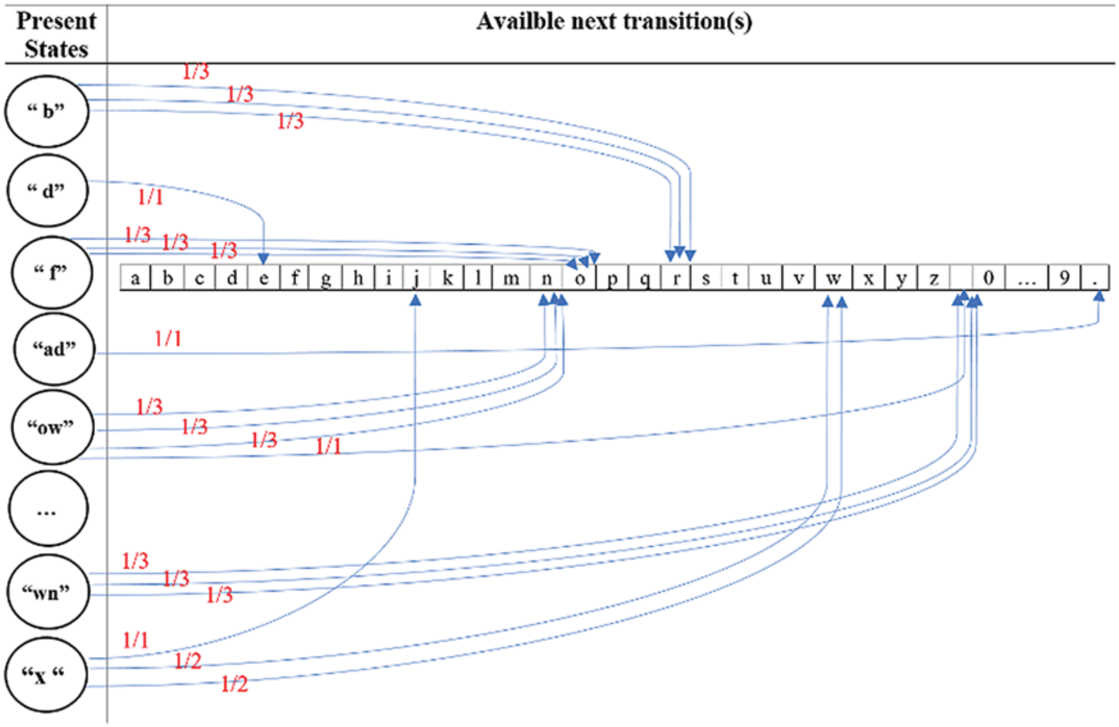
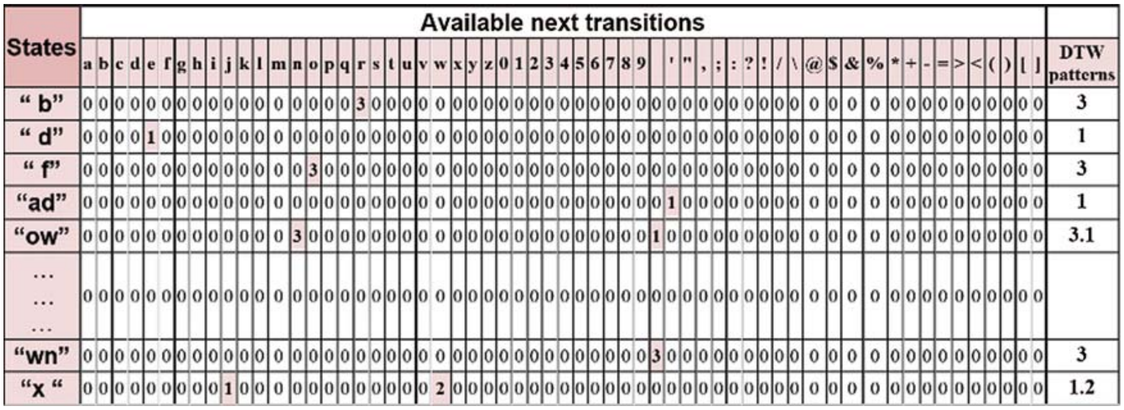


Figure 2: English-text samples representation based RTADZWA

Fig. 3 illustrates the analysis results of the given English sample and represents each state and their transitions based on the second level and alphanumeric approach of the Markov-model.

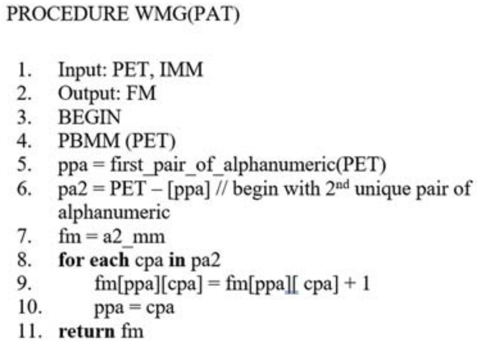


States	Available next transitions																																																											
	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	0	1	2	3	4	5	6	7	8	9		'	"	,	;	:	?	!	/	\	@	$	&	%	*	+	-	=	>	<	(	)	[	]	DTW patterns
" b"	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
" d"	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
" f"	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
"ad"	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
"ow"	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3.1
… … …	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
"wn"	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
"x "	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.2

Figure 3: English text analysis and watermark generation based RTADZWA

The algorithm of text analysis and watermark generation procedure is formally introduced and performed as illustrated in Algorithm 2.

Algorithm 2: Watermark generation algorithm of RTADZWA



```
PROCEDURE WMG(PAT)

Input: PET, IMM
Output: FM
BEGIN
PBMM (PET)
ppa = first_pair_of_alphanumeric(PET)
pa2 = PET – [ppa] // begin with 2nd unique pair of
alphanumeric
fm = a2_mm
for each cpa in pa2
    fm[ppa][cpa] = fm[ppa][ cpa] + 1
    ppa = cpa
return fm
```

Where ppa: previous unique pair of alphanumeric, cpa: current unique pair of alphanumeric.

3.1.3 Watermark Embedding Procedure

Watermark embedding has taken place logically in this method without needing to change the original text. In fact, the feature extraction of the given English-text, watermark key is embedded logically by identifying all non-zero values in the Markov chain matrix. All these non-zero values are sequentially concatenated to form the original pattern of watermark key WMP_O, as defined in Eq. (2) and Fig. 4.

$$a2\_WMP_O \,\&= a2\_mm[ps]\,[ns],\ for\ i\ ,j\ =\ non-zeros\ values\ resulted\ in\ a2\_Markov\_matrix \tag{2}$$

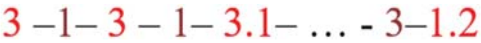
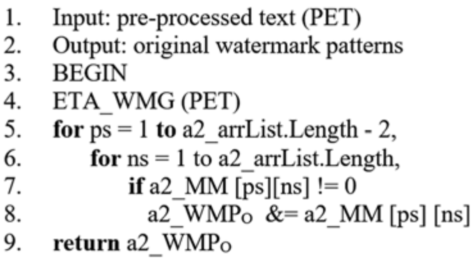

3 –1– 3 – 1– 3.1– … - 3–1.2

Figure 4: The generated original pattern of watermark key WMP_O using RTADZWA

The algorithm of the watermark embedding procedure using the RTADZWA approach is introduced formally and implemented as shown in Algorithm 3.

Algorithm 3: Algorithm of watermark embedding using RTADZWA



```
Input: pre-processed text (PET)
Output: original watermark patterns
BEGIN
ETA_WMG (PET)
for ps = 1 to a2_arrList.Length - 2,
    for ns = 1 to a2_arrList.Length,
        if a2_MM [ps][ns] != 0
            a2_WMP$_O$ &= a2_MM [ps] [ns]
return a2_WMP$_O$
```

Where a2_WMP_O is an original watermark pattern.

3.2 *Watermark Extracting and Detecting Procedures*

This procedure consists of two key algorithms that are extracting and detecting the watermark. However, a2_EWM_A extracted from the obtained will be extracted (AET$_{P}$) and matched by the detection algorithm with a2 WMP_O. AET$_{P}$ is required as input to run this algorithm. Hence, it is necessary to perform the algorithm of watermark generation for obtaining the pattern of watermark for AET$_{P}$ as presented in Fig. 5.

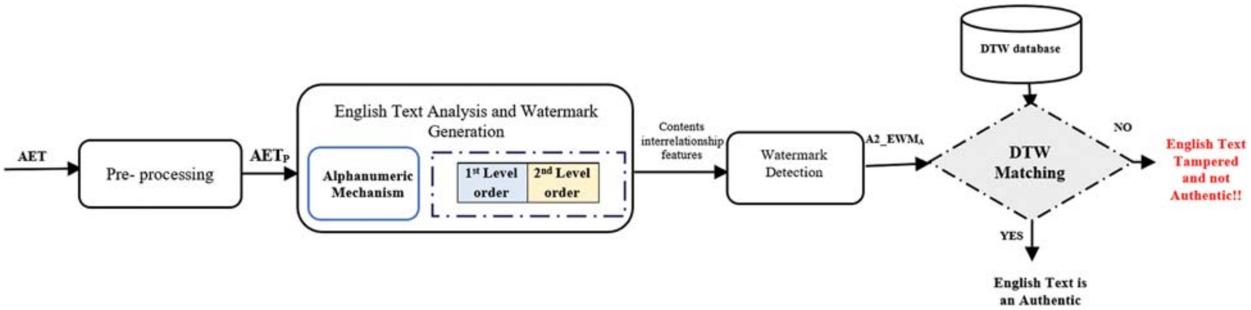
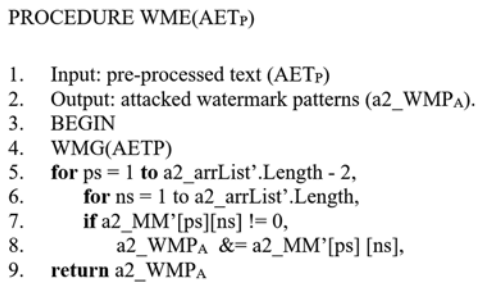

AET
Pre- processing
AETP
English Text Analysis and Watermark Generation
Alphanumeric Mechanism
1st Level order
2nd Level order
Contents interrelationship features
Watermark Detection
A2_EWMA
DTW database
DTW Matching
NO
English Text Tampered and not Authentic!!
YES
English Text is an Authentic

Figure 5: Zero-watermark of RTADZWA procedures of extraction and detection

3.2.1 *Watermark Extraction Procedure*

AET_P should be provided as input to run this algorithm. Though, a2_WMP$_{A}$ is a core output of this algorithm as presented in Algorithm 4.

Algorithm 4: Algorithm of watermark extraction based RTADZWA



```
PROCEDURE WME(AETP)

Input: pre-processed text (AETP)
Output: attacked watermark patterns (a2_WMPA).
BEGIN
WMG(AETP)
for ps = 1 to a2_arrList'.Length - 2,
    for ns = 1 to a2_arrList'.Length,
    if a2_MM'[ps][ns] != 0,
        a2_WMPA &= a2_MM'[ps] [ns],
return a2_WMPA
```

where AET$_{P}$: pre-processed English-text attacked, a2_EWM$_{A}$: attacked pattern of watermark key.

3.2.2 *Algorithm of Watermark Detecting*

a2_WMP$_{A}$ and a2_WMP$_{O}$ should be provided as the inputs needed for this algorithm to run. However, the status of the given English-text is a core output of this algorithm which can be actual or tampered with. The watermark detection process is performed by two sub-steps which are:

- *Main matching* for a2 WMP$_{O}$ and a2 EWM$_{A}$ is achieved. If these two watermark patterns are similar in appearance, then there'll be a warning, "English text contents is authentic and no tampering occurred". Likewise, the note will be rendered "This English text document is tampered and not authentic", and then it continues to the next step.

- *Secondary matching* is performed by matching each state's transition status in the entire produced pattern of watermarks. This means a2_EWM$_A$ of each state is contrasted with an analogous transition of a2_WMP$_O$ as given by Eqs. (3) and (4) below

$$a2\_PMR_T(i,j) = \left| \frac{a2\_WMP_O[i][j] - (a2\_WMP_O[i][j] - a2\_EWM_A[i][j])}{a2\_WMP_O[i][j]} \right|, \quad \textit{for all i,j states and transitions} \tag{3}$$

where a2_PMR_T represents tampering detection accuracy rate value in transition level, (0 <a2_-PMR$_T$ <=1)

$$a2\_PMR_S(i) = \left| \frac{\sum_{j=1}^{n-2} (a2\_PMR_T(i,j))}{total\ StatePatternCount(i)} \right| \textit{for all i} \tag{4}$$

where a2_PMR$_S$: value of tampering detection accuracy rate in state level, (0 < a2_PMR$_S$ <=100).

The weight of every state in the Markov matrix must be determined following the equivalent rate of every state, as is seen in Eq. (5).

$$a2\_Sw = \left| \frac{a2\_PMR_S(i) * Transitions\,frequency(i)}{total\,number\,of\,transitions} \right| \tag{5}$$

where a2_PMR$_S$: is the total matching value in the ith state level.

The ultimate a2_PMR of AET_P and AET are computed by Eq. (6).

$$a2\_PMR = \left| \frac{\sum_{i=1}^{n-2} a2\_PMR_S(i)}{N} \right| \tag{6}$$

The distortion rate of the watermark is the sum of manipulative attacks on the contents of the English context that have been defined by a2_WDR and calculated by Eq. (7).

$$a2\_WDR = 1 - a2\_PMR * 100 \tag{7}$$

The algorithm of watermark detection is formally introduced and applied as seen in Algorithm 5.

The effects of the method of watermark extraction and detection is illustrated in Fig. 6.

4 Implementation and Simulation

A variety of implementation and simulation simulations are conducted to test the accuracy of RTADZWA output and tampering detection. This section outlines a setting for implementation and experimentation, conditions for experiments, typical dataset experimental scenarios, and discussion.

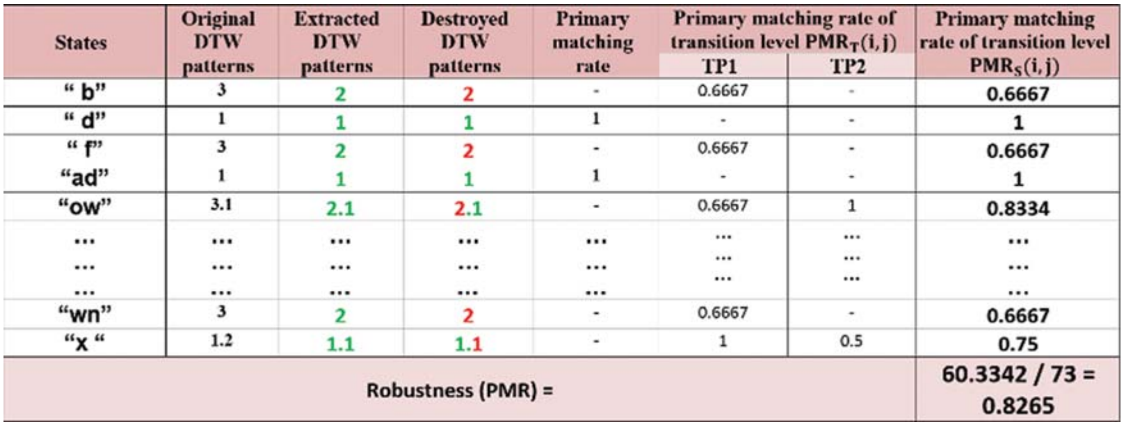
4.1 Simulation and Implementation Environment

The self-developed software was developed to evaluate and assess the efficiency of RTADZWA. The RTADZWA implementing environment is: CPU: Intel Core i7-4650U/2.3 GHz, RAM: 8.0 GB, Windows 10–64 bit, PHP VS Code IDE programming language.

Algorithm 5: Algorithm of watermark detection based RTADZWA

```
PROCEDURE WMD ($a2\_WMP_O$, $a2\_WMP_A$)

Input: pre-processed text ($a2\_WMP_O$, $a2\_WMP_A$)
Output: a2_PMR, a2_WDR
BEGIN
WMG ($a2\_WMP_O$)
WME ($a2\_WMP_A$)
IF $a2\_WMP_A = a2\_WMP_O$
    Print "English document is authentic, and no tampering occurred"
    a2_PMR = 100
Else
    Print "English document is not authentic, and tampering occurred"
for i = 1 to a2_arrList'.Length - 2,
    for j = 1 to a2_arrList'.Length
        IF $a2\_WMP_O[i][j]$ != 0
            pattern Count  +=1
            $a2\_PMR_T(i,j) = \left|\frac{a2\_WMP_O[i][j] - (a2\_WMP_O[i][j] - a2\_WMP_A[i][j])}{a2\_WMP_O[i][j]}\right|$
            transPMRTotal += $a2\_PMR_T$
        Else
            IF $a2\_WMP_A[i][j]$ != 0
            patternCount += $a2\_WMP_A[i][j]$
            $a2\_PMR_S(i) = \left|\frac{\sum_{j=1}^{n-2}(a2\_PMR_T(i,j))}{Total\ StatePatternCount(i)}\right|$
            sWeight = $\frac{a2\_PMR_S(i) * Transitions\ frequency(i)}{total\ no\ of\ transitions}$
a2_SW += stateWeight
a2_PMR = $\frac{\sum_{i=1}^{n-2}(a2\_SW) * Total\ number\ of\ transitions}{Total\ number\ of\ transitions} * 100$
a2_WDR = 1 – a2_PMR * 100
return a2_PMR, a2_WDR
```



States	Original DTW patterns	Extracted DTW patterns	Destroyed DTW patterns	Primary matching rate	Primary matching rate of transition level $PMR_T(i,j)$		Primary matching rate of transition level $PMR_S(i,j)$
					TP1	TP2	
" b"	3	2	2	-	0.6667	-	**0.6667**
" d"	1	1	1	1	-	-	**1**
" f"	3	2	2	-	0.6667	-	**0.6667**
"ad"	1	1	1	1	-	-	**1**
"ow"	3.1	2.1	2.1	-	0.6667	1	**0.8334**
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
"wn"	3	2	2	-	0.6667	-	**0.6667**
"x "	1.2	1.1	1.1	-	1	0.5	**0.75**
Robustness (PMR) =							**60.3342 / 73 = 0.8265**

Figure 6: Results of extraction of watermarks and detection using RTADZWA

4.2 RTADZWA Simulation and Experiment Findings

The performance of RTADZWA refers to the accuracy rate of tampering detection of illegal attacks.

4.2.1 RTADZWA Experiment Under Small (10%) Attack Volumes

Tampering detection accuracy results of RTADZWA under 10% of attack volume of all attacks against all dataset sizes are graphically illustrated in Fig. 7. These results are discussed below.

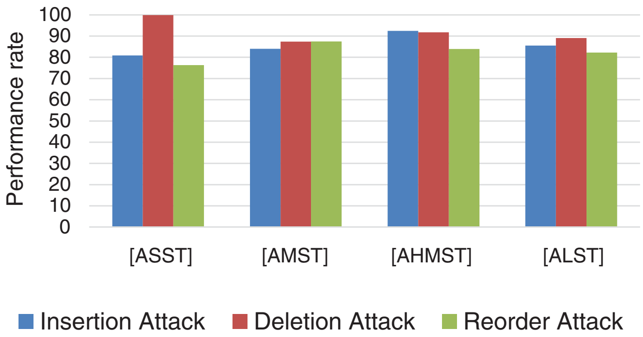


Figure 7: Performance evaluation of RTADZWA under 10% volume of all attacks

From Fig. 7 above, results under 10% attack volume show the best tampering detection rate in all scenarios of deletion attack. This means RTADZWA very sensitive to the small volume of deletion attack.

4.2.2 RTADZWA Experiment Under Mid (20%) Attack Volumes

As observed from the results shown in Fig. 8 under 20% attack volume, RTADZWA gives the best performance in all scenarios of deletion attack, as well as results shown in scenario of 10% attack volumes.

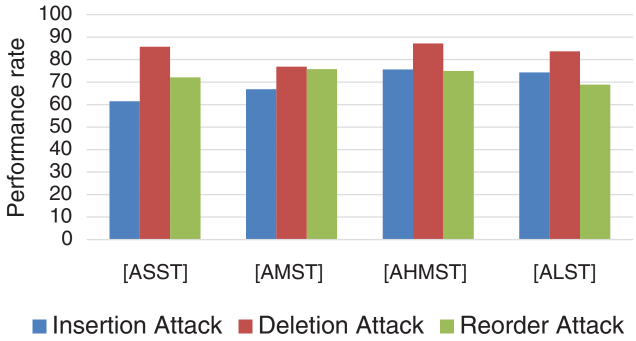


Figure 8: Performance evaluation of RTADZWA under 20% volume of all attacks

4.2.3 RTADZWA Experiment Under Mid (50%) Attack Volumes

As observed from the results shown in Fig. 9 under 50% attack volume, RTADZWA gives the best performance in all scenarios of deletion attack in cases of a very small and middle datasets.

However, in the case of the small and large datasets, the RTADZWA is more sensitive under reorder attacks.

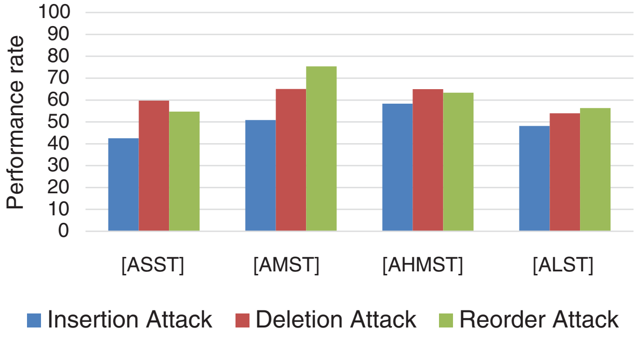


Figure 9: Performance evaluation of RTADZWA under 50% volume of all attacks

4.2.4 RTADZWA Simulation and Experiment Findings Under All Attack Volumes

The performance of RTADZWA refers to the accuracy rate of tampering detection of illegal attacks.

To evaluate the performance of RTADZWA, Scenarios of many studies are performed as shown in Tab. 1, for all forms of attacks and their volumes.

Table 1: Assessment performance of RTADZWA under all volumes

Attack volume (%)	Insertion	Deletion	Reorder
5	93.25	94.79	96.86
10	85.88	89.93	95.32
20	79.13	81.07	91.87
s50	62.30	56.69	88.72

The results shown in Tab. 1 and Fig. 10, it seems that the RTADZWA 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 tampering based on attack types, the results show that the most sensitive tampering detection in all attack volume scenarios is the insertion attack.

5 Comparison and Result Discussion

5.1 Baseline Approaches

The performance results are critically analysed and compared between RTADZWA and baseline approaches UZWAMW and HNLPZWA and show discussion of their effect under the major factors i.e., dataset size, attack types, and volumes to find which approach gives the best performance. Baseline approaches and their objectives are stated in Tab. 2.

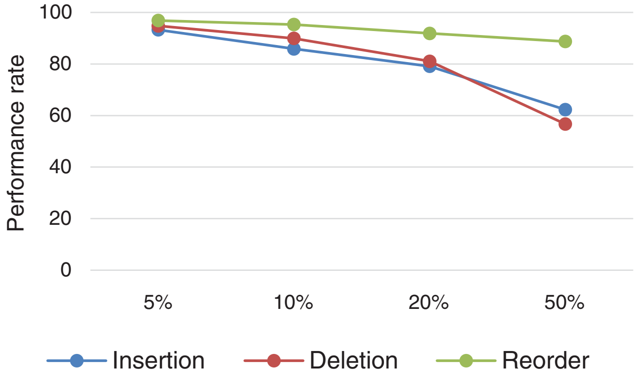


Figure 10: RTADZWA performance under all volumes of various attacks

Table 2: Compared baseline approaches

Approach	Tampering nature and locations	Attack types	Attacks volumes	Objectives
HNLPZWA	Random and multiple	Insertion, deletion, and reorder	5%, 10%, 20% and 50%	Content authentication and tampering detection
UZWAMW				

5.2 Comparative Results

5.2.1 Results of Dataset Impact

This section tests the various data set size impact on watermark reliability against all forms of attacks within their multiple volumes. Tab. 3 shows a comparison of that effect using RTADZWA with HNLPZWA and UZWAMW approaches.

Table 3: Detection accuracy comparison based on the English text size

Dataset size	HNLPZWA	UZWAMW	RTADZWA
[ASST]	67.27	69.53	84.12
[AMST]	63.80	68.13	84.38
[AHMST]	59.23	65.11	85.12
[ALST]	54.47	62.07	85.39

The comparative results as shown in Tab. 3 and Fig. 11 reflects the performance of RTADZWA approach. The results show that in the proposed RTADZWA approach, the highest effects of dataset size that lead to the best performance are ordered as ASST, ALST, AMST, and AHMST, respectively. This means that performance increased with increasing text length and decreased with decreasing text length. On the other hand, results show that RTADZWA approach outperforms both HNLPZWA and UZWAMW approaches in terms of watermark robustness under all scenarios of dataset sizes.

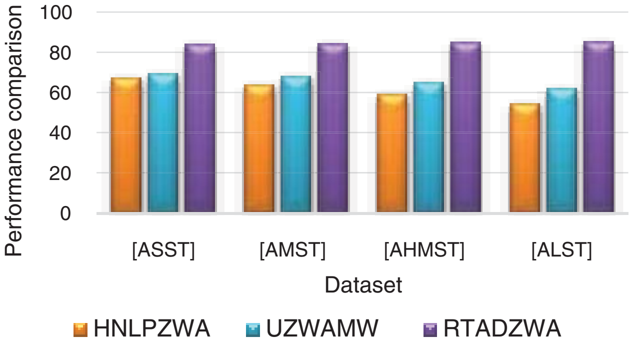


Figure 11: English text size-based comparison of tampering detection impact

5.2.2 Results of Attack Type Impact

Tab. 4 shows a comparison of the different attack type's effect on tampering detection accuracy of RTADZWA, HNLPZWA, and UZWAMW approaches against all dataset scales and all attack volume scenarios. In all cases of attack types, low effect detected under insertion attack by RTADZWA and baseline HNLPZWA and UZWAMW approaches because deletion and reorder attacks represent both insertion and deletion tampering at the same time.

Table 4: Detection impact comparison based on attack type

Method	Insertion	Deletion	Reorder
HNLPZWA	71.28	59.99	37.23
UZWAMW	80.02	66.25	44.88
RTADZWA	80.14	80.62	93.19

In general, the comparative results shown in Tab. 4 and illustrated in Fig. 12 show that RTADZWA outperforms baseline HNLPZWA and UZWAMW approaches with high-performance rate and low effect of attack types. This means that the proposed RTADZWA approach is strongly recommended and applicable for content authentication and tampering detection of English text under all attack types.

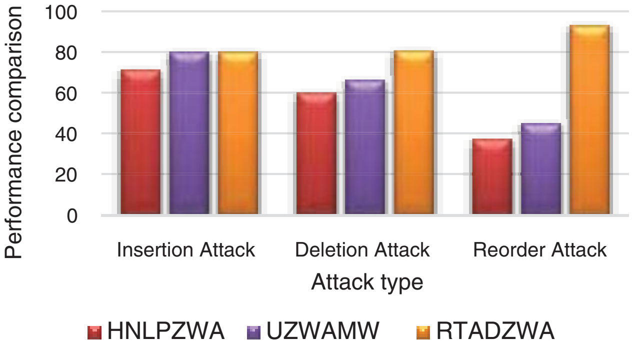


Figure 12: Attack type-based comparison of tampering detection effect

5.2.3 Results of Attack Rates Impact

Tab. 5 provides a comparison of the multiple attack volume effects on the performance of tampering detection for both dataset size, and volume scenarios. The comparison is performed using RTADZWA, HNLPZWA, and UZWAMW approaches.

Table 5: Detection accuracy comparison based on attack rates

Attack volume (%)	HNLPZWA	UZWAMW	RTADZWA
5	82.09	83.60	94.32
10	72.74	74.33	91.06
20	57.71	59.39	82.14
50	13.66	37.56	68.96

Tab. 5 and Fig. 13 show how the performance is influenced by low, mid, and high attack volumes. In cases of mid and high attack volumes, a very high effect is detected by baseline HNLPZWA and UZWAMW approaches. However, a very low effect is detected by the proposed RTADZWA approach. In Fig. 11, it can be seen that if the attack volume increases, the tampering detection accuracy also increases. In all cases of low, mid, and high attack volumes, it seen RTADZWA outperforms baseline HNLPZWA and UZWAMW in terms of performance in all scenarios of low, mid, and high volumes of all attacks. This means that RTADZWA approach is strongly recommended for content authentication and tampering detection of English text under all volumes of all attacks.

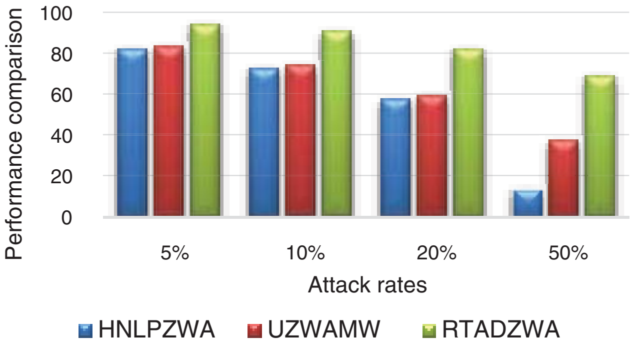


Figure 13: Attack rate-based comparison of tampering detection accuracy

6 Conclusion

Based on second level order and alphanumeric mechanism of hidden Markov model, a novel hybrid approach of natural language processing and zero-watermarking has been developed which is abbreviated as RTADZWA. The core aim of the proposed RTADZWA is content authentication and tampering detection of English text transmitted via the Internet. RTADZWA approach is implemented in PHP programming language using VS code IDE. The simulation and experiments are performed on various standard datasets under different volumes of insertion, deletion, and reorder attacks. RTADZWA approach has been compared with HNLPZWA and UZWAMW approaches. Comparison results show that RTADZWA outperforms baseline HNLPZWA and UZWAMW approaches in terms of general performance which represents watermark capacity, watermark robustness and tampering detection accuracy under all scenarios of all attack types and volumes. For future work, the author will intend to improve the performance using the high-level of the alphanumeric mechanism of the Markov model.

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.2/25/42), 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. Abd-Eldayem, "A proposed security technique based on watermarking and encryption for digital imaging and communications in medicine," *Egyptian Informatics Journal*, vol. 14, pp. 1–13, 2013.

[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] S. G. Rizzo, F. Bertini and D. Montesi, "Fine-grain watermarking for intellectual property protection," *EURASIP Journal on Information Security*, vol. 10, 2019. https://doi.org/10.1186/s13635-019-0094-2.

[5] 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, pp. 1–15, 2020.

[6] 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.

[7] P. Selvama, S. Balachandran, S. Pitchai and R. Jayabal, "Hybrid transform based reversible watermarking technique for medical images in telemedicine applications," *ELSEVIER Optik*, vol. 145, pp. 655–671, 2017.

[8] N. Hurrah, A. Parah, N. Loan, A. Sheikh, M. Elhoseny *et al.*, "Dual watermarking framework for privacy protection and content authentication of multimedia," *ELSEVIER Future Generation Computer Systems*, vol. 94, pp. 654–673, 2019.

[9] A. Panah, R. Van, T. Sellis and E. Bertino, "On the properties of non-media digital watermarking: A review of state-of-the-art techniques," *IEEE Access*, vol. 4, pp. 2670–2704, 2016.

[10] C. Qin, C. Chang and T. Hsu, "Fragile watermarking for image authentication with high-quality recovery capability," *KSII Transactions on Internet and Information Systems*, vol. 7, no. 11, pp. 2941–2956, 2013.

[11] S. Parah, J. Sheikh and G. Bhat, "Stegnmark: A joint stego-watermark approach for early tamper detection," *Springer International Publishing Switzerland*, vol. 660, pp. 427–452, 2017.

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

[13] M. Taleby, Q. Li, X. Zhu, M. Alazab and J. Zhang, "A novel intelligent text watermarking technique for forensic identification of information on social media," *Computers and Security*, vol. 90, pp. 1–14, 2020.

[14] S. Parah, J. Sheikh, J. Akhoon and N. Loan, "Electronic health record hiding in images for smart city applications: A computationally efficient and reversible information hiding technique for secure communication," *ELSEVIER Future Generation Computer Systems*, vol. 108, pp. 935–949, 2020.

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

[16] 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.

[17] R. Alotaibi and L. Elrefaei, "Improved capacity 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.

[18] 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.

[19] 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, pp. 381–392, 2014.

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

[21] A. Rahma, W. Bhaya and D. Al-Nasrawi, "Text steganography based on unicode of characters in multilingual," *Journal of Engineering Research and Applications*, vol. 3, no. 4, pp. 1153–1165, 2013.

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

[23] M. Bashardoost, M. Rahim, T. Saba and A. Rehman, "Replacement attack: A new zero text watermarking attack," *3D Research*, vol. 8, no. 1, 2017. https://doi.org/10.1007/s13319-017-0118-y.

[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, W. Song, A. Li, Y. Zhang and R. Tao, "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] Z. Ali, M. Shamim, G. Muhammad and M. Aslam, "New zero-watermarking algorithm using hurst exponent for protection of privacy in telemedicine," *IEEE Access*, vol. 6, pp. 7930–7940, 2018.

[27] 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.

[28] M. Ghilan, F. Ba-Alwi and F. N. Al-Wesabi, "Combined markov model and zero watermarking to enhance authentication of arabic text," *Journal of Computational Linguistics Research*, vol. 5, no. 1, pp. 26–42, 2014.

[29] 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.

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