Journal of Digital Imaging (2020) 33:555–562
https://doi.org/10.1007/s10278-019-00296-y

ORIGINAL PAPER

Automated Misspelling Detection and Correction in Persian Clinical Text

Azita Yazdani[1,2] · Marjan Ghazisaeedi[3] · Nasrin Ahmadinejad[4,5] · Masoumeh Giti[5,6] · Habibe Amjadi[7] · Azin Nahvijou[8]

Published online: 10 December 2019

Abstract

Accurate electronic health records are important for clinical care, research, and patient safety assurance. Correction of misspelled words is required to ensure the correct interpretation of medical records. In the Persian language, the lack of automated misspelling detection and correction system is evident in the medicine and health care. In this article, we describe the development of an automated misspelling detection and correction system for radiology and ultrasound's free texts in the Persian language. To achieve our goal, we used n-gram language model and three different types of free texts related to abdominal and pelvic ultrasound, head and neck ultrasound, and breast ultrasound reports. Our system achieved the detection performance of up to 90.29% for radiology and ultrasound's free texts with the correction accuracy of 88.56%. Results indicated that high-quality spelling correction is possible in clinical reports. The system also achieved significant savings during the documentation process and final approval of the reports in the imaging department.

Keywords Natural language processing · Spelling correction · N-gram language model · Ultrasound · Radiology reporting

Abbreviations

HIS	Hospital information system
EPR	Electronic patient record
EHR	Electronic health record
OCR	Optical character recognition

Introduction

Documentation process is an important part of EHR. Clinical reports that have special place in the documentation are stored in narrative text and structured data. A text is an important type of data in the field of biomedicine. For clinicians, the written text of medical findings is still the basis for decision making [1]. Time and efficiency pressures have ensured clinicians' continued preference for unstructured text over entering data in structured forms when composing progress notes [2]. Accurate clinical documentation is critical in health care quality and safety [3]. Misspellings take place since clinical texts are written under time pressure [4]. Several studies have been conducted to check the misspellings in clinical notes. According to the survey, the percentage of misspells is from

✉ Azin Nahvijou
aznahvi@yahoo.com

1 Department of Health Information Technology, School of Management and Information Sciences, Shiraz University of Medical Sciences, Shiraz, Iran

2 Health Human Resources Research Center, School of Management and Information Sciences, Shiraz University of Medical Sciences, Shiraz, Iran

3 Department of Health Information Management, School of Allied Medical Sciences, Tehran University of Medical Sciences(TUMS), Tehran, Iran

4 Advanced Diagnostic and Interventional Radiology Research Center (ADIR), Tehran University of Medical Sciences (TUMS), Tehran, Iran

5 Medical Imaging Center, Cancer Research Institute, Imam Khomeini Hospital, Tehran, Iran

6 Breast Cancer Research Center, Cancer Institute, Tehran University of Medical Sciences (TUMS), Tehran, Iran

7 Human Resource Department, Imam Hospital, Tehran University of Medical Sciences (TUMS), Tehran, Iran

8 Cancer Research Center, Cancer Institute of Iran, Tehran University of Medical Sciences(TUMS), Tehran, Iran

1 to 10% [5]. For example, the rate of misspelling in the French clinical text is about 10%, in Australian clinical notes is about 2.3%, in Swedish clinical corpora is 1.1%, in English clinical notes is 2%, and in Stockholm EPR corpus is 7.6%, and also, the misspell rate in the follow-up reports is above 10% [6]. According to the results of a study [7] that was conducted on the Mayo's clinic radiology reports, the rate of error in chest X-ray reports was 3.2% and in the neuroradiology reports was 19.7%. A study conducted in England in 2015 showed that 23% of CT reports and 32% of MRI reports had at least one misspelled word [8]. Although misspellings may not put a significant cognitive pressure on readers, they continue to challenge the use and processing of unstructured texts or free text and the effectiveness of automated systems, such as text mining and information retrieval, information extraction, text summarization, and encoding [9, 10]. Therefore, detection and correction of misspelled words are necessary for effective use of clinical notes and texts. Hospital users need various support tools such as spelling correction tool to enter data in EPR. With the emergence of automated spell correction systems, an opportunity has been created for these systems to be used as a pre-processing step [6] in various text processing applications, such as information retrieval and text categorization [11]. In fact, these systems are considered as a process to improve the text mining [12]. In the medical domain, spelling correction is used to expand acronyms and abbreviations, truncation, and correction of misspelling. According to various studies, these cases reach 30% in clinical content [13]. In the past two decades, spelling correction techniques for clinical texts have been widely studied [2]. Most of these studies have focused on EHR [14], and some studies have focused on consumer-generated texts in the field of health care [15, 16]. Ruch and colleagues introduced a French clinical record spell checker, which corrects up to 95% of misspells in the text [17]. In order to correct misspellings in Hungarian clinical text, a context-aware system was introduced by Siklósi and colleagues based on statistical machine translation with the accuracy of 87.23% [18]. Grigonyte and colleagues developed a system for correcting misspelling in the Swedish clinical text with 83.9% accuracy and 76.2% recall [19]. Based on the Google spell checker, Zhou and colleagues developed a spelling correction system that can accurately correct 86% of the typographic and linguistic errors in the daily medical vocabularies [16]. In the study [20], the existing spelling correction system has been cited in the vaccine safety reports. Recall and precision in this system were 74% and 47%, respectively. Wong and colleagues developed a system with an accuracy of 88.73% in order to correct misspellings in clinical reports in the real time. This system uses semantic statistical analysis in web data to automatically correct misspells [2]. Doan and colleagues introduced a system to correct the misspelling of drug names based on the Aspell algorithm. Precision in this system has been reported at 80% [21]. Lai and colleagues' article is among the recent articles that proposed auto spell correction in the field of medicine. The spell-checking system provided by them is based on a noisy channel model, with misspelling correction accuracy of 80% in clinical texts [22]. Other recent articles in this area include the studies of [23, 24]. In the study [23], the authors presented an unsupervised context-sensitive spelling correction method for English and Dutch clinical free texts. With the advent of OCR technology in the last two decades and its use in the medical field; many spell correction systems have been developed to automatically detect and correct OCR errors. One of these medical correction systems is the work of [25] that detects and corrects misspells in French clinical texts.

Two spell checkers have been developed for the Persian language including Virastyar and Vafa spell-checker, both of which are focused on detection and correction of misspells in general fields. The linguistic models in both spell checkers have been trained and tested in general cases (news, political, sport, economic, etc.). Therefore, there is a need for Persian spell checker in specialized areas, including health care. In this study, an automated spelling detection and correction system was developed in the Persian clinical reports in the department of imaging at Imam Khomeini hospital.

Material and Methods

Our non-word spell checker is based on the n-gram language model and is presented in the orthography of words in Persian clinical reports. A non-word error is a misspelled word that is caused by a typographical error [32]. This article uses two dictionaries: dictionary of specialized vocabulary and general vocabulary dictionary. The dictionary of specialized vocabulary used in this article is based on translation of various lexical resources such as notes of breast ultrasound, head and neck ultrasonography, and abdominal and pelvic ultrasound in Persian. This dictionary is used to detect misspelling of medical terms in medical reports. In fact, the medical words and terms in various medical texts are extracted and used as a dictionary. Also, in order to detect misspelling in general words, the dictionary of Vafa spell-checker, a comprehensive dictionary of Persian words, is used [26]. We trained and tested our spell checker on various types of reports collected from the hospital information system of the department of imaging at Imam Khomeini Hospital in Tehran. In order to evaluate the spell checker developed in this paper, the accuracy, precision, F-measure, and recall measurements were used.

Misspelling Detection

The problem of spelling correction can be divided into two parts: error detection and error correction [20]. We used dictionaries to detect misspells. Any word that does not exist in

dictionaries is recognized as misspelling. Based on Eq. (1), all the words in a sentence are evaluated by a dictionary one by one, and if a word is not in the dictionary, it is detected as misspell w_n and the rest of the words are divided into two sequences. This is the sequence of words before the misspell $w_{n-m} \ldots w_{n-1}$ and the sequence of words after the misspell $w_{n+1} \ldots w_{n+m}$. These two sequences are used in language models such as n-gram to correct the misspell w_n.

$$\underbrace{\{W_{n-m} \ldots W_{n-2} \; W_{n-1}}_{\text{Previous words}} \; \overset{\text{Error}}{\longleftrightarrow} \; W_n \; \underbrace{W_{n+1} \; W_{n+2} \ldots W_{n+m}\}}_{\text{Subsequent words}} \tag{1}$$

Medical texts contain specialized vocabulary in the field of medicine along with the general words. Our spell checker uses a comprehensive dictionary to detect misspells. This dictionary has two sections of general and specialized words. The Vafa spell-checker dictionary is used for the general section, which is a spell checker for the Persian language. The dictionary contains 1,095,959 words, all of which are general words and lacking the specialized terms in the field of medicine (Table 1).

In this article, the trained texts were used to create a custom dictionary. This dictionary uses specialized vocabulary in breast ultrasonography, head and neck ultrasonography, and abdominal and pelvic ultrasonography texts, which was also combined with the translation of the Radiological Sciences Dictionary by David J Dowsett [27], to identify misspelling of specialized words. The dictionary contains 10,332 words, which are all specialized words in the field of breast ultrasound, head and neck ultrasound, and abdominal and pelvic ultrasound. This specialized dictionary lacks general words. The extensive dictionary was compared with the dictionary of Radiological Sciences Dictionary with the help of a custom program produced by the researchers of this article in order to prevent specialized words from being included more than once in the dictionary (some words may be in two dictionaries).

Table 1 Sample of the general dictionary

ID	Word	Frequency	POS_tag	POS
1	آبادها ان	49	6	Plural noun + personal pronouns
2	آبادها تان	49	6	Plural noun + personal pronouns
3	آبادم	49	6	Plural noun + personal pronouns
4	آبادت	2	3	Plural noun + personal pronouns
5	آباد	2	3	Plural noun + personal pronouns
6	آبادمان	3	3	Plural noun + personal pronouns
7	آبادتان	1	3	Plural noun + personal pronouns

Misspelling Correction

When a misspell is detected, the system should offer a list of suggestions that can replace the misspelled word. In this paper, n-gram language model was used for spelling correction. N-gram can dictate suggestions based on orthographic and edit distances. Edit distance means the number of incorrect characters in the misspelling detection. Almost 80% of errors are within edit distance 1 [32]. Based on Damerau-Levenshtein edit distance, minimal edit distance between two strings, where edits are the following:

- Insertion: Insert additional characters in the word.
- Deletion: Remove characters from the word.
- Substitution: Replacing the wrong character with the correct one.
- Transposition: Displacement of two characters with the correct word [28].

We used Damerau-Levenshtein distance to generate suggestion lists. Dictionary is used to generate suggestion lists. Suggestion lists are produced according to edit distance and the similarity of words in the dictionary with the wrong word. Sometimes there are many suggestions that make it difficult for the user to select the correct word, so there is a need for a scoring method to limit the list of suggestions and introduce the most likely candidates in the list. For this purpose, the n-gram language models have been used in this paper.

N-Gram Language Model

N-gram is a method of examining the n sequence of an item in a text or audio [29]. These items can be phoneme, syllable, letters, words, and base pairs according to the application. N-grams are generally collected from speech or text corpus. N-grams (Eq. 2) are used to predict the next word in a sequence, a probabilistic predictor model that calculates the probability of a word occurring after a sequence of n–1 words based on the Markov chain model [30].

$$\begin{aligned} P(w) &= P(w_1 w_2 \ldots w_n) \approx \prod_{i=1}^{n} p(w_i \mid w_{i-k} \ldots w_{i-1}) \\ &\approx \prod_{i=1}^{n} P(w_1) P(w_2)\, P(w_3) \ldots \; p(w_n) \quad \textit{Equation 2 : N-gram.} \end{aligned} \tag{2}$$

N-grams are named according to their size. Unigram is size one, bigram is size two, and trigram is size 3 (Eq. 3). The larger ones are called four-gram and so on.

$$\begin{array}{lll} W_n & W_{n+1} & W_{n+2} \\ \text{Trigram} & \longleftrightarrow & \\ \text{Bigram} & \longleftrightarrow & \\ \text{Unigram} & \longleftrightarrow & \end{array} \tag{3}$$

The most popular n-gram models are the bigram and trigram models.

$$P(w)\approx\prod_{i=1}^{n}p(W_i|W_{i-1})\ \textit{Equation 4 : Bigram equation.} \quad (4)$$

The bigram (Eq. 4), trigram (Eq. 5), and four-gram linguistic models are used in this paper for misspelling correction.

$$P(w)\approx\prod_{i=1}^{n}p(W_i|W_{i-2},W_{i-1})\ \textit{Equation 5 : Trigram equation} \quad (5)$$

After identifying the wrong word and producing candidates based on Damerau-Levenshtein distance, the best candidate should be determined among the candidates. Our spell checker uses edit distance 2 to generate candidates, which causes the number of candidates to be greater than the edit distance 1.

$$\underbrace{\{W_{n-m}\ \ldots\ W_{n-2}\ W_{n-1}}_{\text{Left Pairs}}\ \overbrace{C_i}^{\text{Candidates}}\ \underbrace{W_{n+1}\ W_{n+2}\ \ldots\ W_{n+m}\}}_{\text{Right Pairs}} \quad (6)$$

In this paper, a combination of bigram, trigram, and four-gram was used to enhance the accuracy of spelling correction process. In previous work, it has been shown that four-gram model is more accurate in the misspelling correction process [10], but there may not be four sequences in the trained texts, so the four-gram model cannot be used to check all sequences. Trigram also has a higher accuracy than bigram, but this language model may not have triple sequences in trained data. In proposed spell checker, we combine all three language models to grant weight to candidates. Based on Eq. (7), the bigram, trigram, and four-gram language models are calculated for all candidates. Since the four-gram language model is more effective in misspelling correction, it will therefore have more weight and a greater impact on the correction of wrong words. Similarly, trigram and bigram are calculated for each individual candidate and are based on their weight gain.

$$\text{Weight}\ (\text{Candidate}_i)=\frac{\sum_{i=1}^{n}\text{Fourgram}(\text{Candidate}_i)^4,\text{Trigram}(\text{Candidate}_i)^3,\text{Bigram}(\text{Candidate}_i)^2,\text{Unigram}(\text{Candidate}_i)^1}{P(\text{Candidate}_i)}\ \textit{Equation 7 : Weighting formula for the produced candidates.} \quad (7)$$

Based on equation 7, the sum of probabilities of unigram, bigram, trigram, and four-gram language models for the candidate *i* are calculated, and the result of this calculation is divided into the probability of candidate's availability in the trained texts. In addition, the power granted to language models depends on their impact on candidate's determination. For example, having 4 sequences in trained texts has a higher value than 3 sequences. Thus, the result obtained from calculating the four-gram probability compared with trigram probability and also the results of trigram probability compared with bigram will be more valuable. Using Eq. (7), each candidate for w_n misspell is considered a weight, and ultimately, the candidate with the highest weight is considered as the correct word.

$$\underbrace{W_{n-m}\ldots W_{n-2}W_{n-1}}_{\text{Previous words}}\ \overbrace{C_{\max\ \text{weight}}}^{\text{Best Candidate}}\ \underbrace{W_{n+1}W_{n+2}\ldots W_{n+m}}_{\text{Subsequent words}} \quad (8)$$

To calculate $P(Candidate_i)$, the number of candidate$_i$ repetitions was devided by all existing words in the train texts.

For example, in the sentence " د ب ا ا ا ت ن د لفن نود با ذا ا ت و م هو د ا ت" availability of all words in the dictionary was evaluated, and since the word " لفن" was not in the dictionary, it was recognized as the wrong word. After identifying the wrong word, candidates are generated in order to replace the wrong word. Damerau-Levenshtein was employed for this purpose (38 candidates were generated). For each of these candidates, the unigram, bigram, trigram, and four-gram were calculated on the basis of collected corpus, respectively (Table 2).

Finally, the most appropriate candidate was selected using Eq. (7) and n-gram calculations in Table 2. Accordingly, the "لنف" candidate (weighing 0.001220) is suggested as the most appropriate choice.

Data

We used three data sources that were provided by the HIS of the department of imaging at Imam Khomeini Hospital in Tehran. In order to produce n-gram language models, the texts were divided into two sets of training set and test set. The training set consisted of three different types of medical reports. The first dataset included reports from patients who

Table 2 Some generated candidates and their n-gram calculated

No	Candidate	Unigram	Bigram	Trigram	Four-gram
1	لن	0.0000514	0.000652	0.001691	0.0000808
2	نف	0.0000171	0.0006499	0.0000675	0
3	لنف	0.0012168	0.0018694	0.0023219	0.0023186
4	فج	0.0007866	0.0001720	0.0000063	0
5	نفن	0.0000220	0.0000025	0	0
6	دلفان	0.0000002	0	0	0

have performed breast ultrasonography to check the presence of lymph nodes in the time period of March 2015–July 2018. The training set consisted of 15,639 reports and 871,275 words. In the training set, 10,442 misspelled words were identified (1.2%). The test set contained 249 reports with 101,863 words. The test set also had 8454 misspelled words (0.83%). The second data set consisted of head and neck ultrasonography reports that contained all texts that had been entered by the medical typist in the HIS since 2014. At this stage, 15,472 reports were compiled with 168,055 words, which also had 1675 misspelled words (1%). The test set contained 75 reports with 26,856 words and 2550 misspellings. The third data set consisted of abdominal and pelvic ultrasound reports recorded between April 2015 and July 2018. Also, 3531 reports were considered as a training set that contained 106,084 words and 1509 misspellings (1.42%). The test set contains 428 reports with 19,264 words and 187 misspellings (0.97%). The error rate in our corpus was 1.15%.

Preprocess

Before processing the texts to standardize letters and spaces, they must be preprocessed. In fact, at this point, all the characters in the text must be replaced by their standard equivalent. In Persian script processing, similar to Arabic script, there are always problems with using equivalent of Arabic characters, including the letters " ", " " and (ۀ أ ؤ إ). The first step is to fix the problems associated with these letters by standardization them. The following inconsistencies were found in the literature collected from the HIS of the department of imaging at Imam Khomeini Hospital:

(a) Different encoding for some characters**:** In Persian script processing, similar to Arabic script, there are always problems with using equivalent of Arabic characters, including the letters " ", " " and (ۀ أ ؤ إ). All these characters were standardized in corpus.

(b) Extra spaces: Across the corpus, there is a great deal of space, half space, and extra tabs. In the pre-processing phase, all of these items were removed.

(c) Variety of text with uppercase and lowercase letter: Sometimes uppercase or lowercase a of letter "آ" is mistakenly used in a word that contains this letter. For example, instead of writing "آبدار" , it has mistakenly been written "ابدار". To solve this problem, the words that begin with the letter "ا" were modified and transformed into the uppercase of the letter "آ". Also, the words that had the letter "ا" among their constituent characters (except the primary character) were modified and changed to its lowercase form.

(d) Different ways to add suffix to main words: The suffixes " ها" " ها" "تن" "ت" etc. are placed at the end of the words in three different forms such as: "من ا ب" "من ا ب ت" "من ا ب ت" . In order to standardize this type of words, all suffixes were added to the end of the words.

(e) Different ways of adding prefixes to main words: The prefixes of "ب" " ب" "د م" "ت" "م" come in three different forms such as: "م ود" "م ود" "م ود". In order to standardize this type of words, all prefixes were added to the beginning of the words using the half-space.

(f) Removing the character "-": This character causes the words to be stretched like: " ب" " ب" converted to "ب".

(g) Different ways of attaching the components of compound words: A compound word is a word that makes a new concept and meaning by combining two or more independently known words. For example, in the word "د پ", each of the components of the word (i.e., the words "پ" and "مد") has a well-known and independent meaning that is not related to the new combination. According to the existing sources of Persian, all of the compound words in the corpus were modified to a unified standard form.

(h) Multiple spelling words: There are words in Persian that contain two or more spelling forms. In these words, the letters that have the same sound and called homophones (ا/ ع ت/ ث/ / ح/ ه ذ/ / /ظ غ/ق) are used. Like"اتاق" and "اق ا", which according to the Academy of Persian Language and Literature, its correct form is "اتاق". Given the multiplicity of these types of words, credible sources were used to identify and standardize these types of words in the corpus.

All preprocessing and language model producing steps were performed using the "normalizer" function of the "normalizer" class in the jhazm library developed in Java.

Misspelling Analysis

We checked all kinds of misspellings on different data sources. Four types of misspellings are common in the texts. The occurrence of these errors in each of the three datasets has been investigated (Table 3). In the breast ultrasound and head and neck ultrasound, deletion is more common than other types of errors. Also, substitution errors are common in breast ultrasound datasets as well as abdominal and pelvic ultrasound datasets. In abdominal and pelvic ultrasound data, insertion errors were the most common types of error, and ultimately, the transposition error in all three sources has least common.

The datasets were also examined for the number of edit distances for spell correction. In all three datasets, 85% misspellings required edit distance 1 to correct the wrong word. Other misspells were also corrected with the edit distance 2. Among all three notes, only abdominal and pelvic ultrasound dataset and head and neck ultrasound had the edit distance 3.

Table 3 Error types in three data sources

Error types	Breast ultrasound (%)	Head and neck ultrasound (%)	Abdominal and pelvic ultrasound (%)
Insertions	23.9	15.1	41.3
Deletions	34.6	39.4	15.3
Substitutions	30.1	31.8	34.2
Transpositions	11.4	13.7	9.2

The edit value required to correct words in all three categories is shown in Table 4.

Results

We evaluated our system's efficiency on misspelling detection and correction. For misspelling detection in the test sets, precision, recall, harmonic mean (F-measure), and accuracy were investigated. In training sets, we only investigated recall, because all the correct words in the training sets were added to the dictionary. In other words, the precision was 100%. Results of the training sets and test sets are shown in Tables 5 and 6, respectively.

In the process of misspelling detection, our system's F-measure range was 88.62 (based on breast ultrasound reports) to 90.29% (on abdominal and pelvic ultrasound reports) in the test set.

Correction accuracy of our system was 81.77 (Breast ultrasound) to 88.56% (head and neck ultrasound). The accuracy range in the training sets was calculated to be from 91.84 to 93.91%. The real-time use of this system not only affects the accuracy of documentation, but also affects the patient's clinical care. The use of the system presented in this article corrected 90.44% of misspelling in radiology and ultrasound reports, and as an intermediate task, it prepared radiology and ultrasound reports for some operations such as text mining and information retrieval.

Table 4 Minimum edit distances needed to transform misspelled words into correct words—test sets

Edit distance	Breast ultrasound (%)	Head and neck ultrasound (%)	Abdominal and pelvic ultrasound (%)
1	81.5	87.9	82.6
2	18.1	11.3	16.3
3+	0.4	0.9	1.1

Table 5 Performance of spell correction on three data sources—training sets

	Breast ultrasound (%)	Head and neck ultrasound (%)	Abdominal and pelvic ultrasound (%)
Recall	93.8	94.81	95.35
Accuracy	92.71	91.84	93.91

Discussion

This system was used as part of the hospital information system. It can also examine free texts before being registered on HIS. It is able to detect and correct spelling errors within clinical reports in two modes. In the settings of this system, you can select the list suggestion mode that the user will participate in the process of selecting the correct word. In automatic mode, the system replaces the incorrect word with the highest probability of candidate based on the n-gram language model calculation (Eq. 7). Our proposed system is successfully able to detect spelling mistakes (detection accuracy up to 90.29%), facilitate rapid report correction (correction accuracy up to 88.56%), and improve clinical reports in the hospital information system. In the Department of Imaging at Imam Khomeini Hospital, radiologists dictate radiology reports, which are typed by medical typists and are returned to the radiologists for editing, and that eventually, the final report is made. Then, the reports will be stored in the HIS. This process takes 30 min in average at best. Activities that do not add value are named as Muda [33]. From the process of typing reports to final registration in the radiology and ultrasound department of imaging at hospital, there is an expected time in which there is no value added activity. By using our developed software, misspelled words are quickly corrected at the time of writing. As a result, the time between writing reports and final confirmation of those reports will be reduced and leads to decreased Muda. Misspelling is common in radiology reports [31]. Frequent interruptions and an increasingly fast work pace contribute to such errors. Errors, depending on their type, have different effects, such as compromising patient health, creating ambiguity, and reducing the credibility of radiologists. So, avoiding error in radiology report is essential

Table 6 Performance of spell correction on three data sources—test sets

	Breast ultrasound (%)	Head and neck ultrasound (%)	Abdominal and pelvic ultrasound (%)
Precision	90.44	89.67	90.12
Recall	87.59	89.35	89.43
F-measure	88.62	90.08	90.29
Accuracy	81.77	88.56	84.14

[31]. The awareness should be raised among radiologists that errors in radiology reports will inevitably occur and the existence of automated misspelling detection and correction systems can, in addition to improving the quality of reports, reduce the time sent to correct these types of errors. Despite the existence of spelling correction systems for the Persian language, these systems have general use and are not specific to medical domain. Therefore, in this paper, we developed a system that can automatically detect and correct misspellings in Persian radiology and ultrasounds. Merging automatic spell checking systems principally in areas that are critical for patient safety such as entry of allergy, medication, diagnosis, and problem has the potential to significantly improve the quality and accuracy of electronic medical records. This system can be installed as an add-in program on Microsoft office word. It can also be used as an API in the HIS system. Our system performed well on all three corpuses on which it was tested. The performance of our system was at its best in all categories, particularly in the head and neck ultrasound, followed by abdominal and pelvic ultrasound, and finally the breast ultrasound reports. We found that rescoring the suggestion list using word frequencies leads to both a notable increase in the precision of misspelling detection and an increase in the correction accuracy. The equation presented in this article (Eq. 7) for weighting the generated candidates can be used in other languages to improve the accuracy of correction of a non-word spelling. The weighting equation presented by combining unigram, bigram, trigram, and four-gram more accurately selects the correct word among the generated candidates. Due to the fact that the n-gram language model is used in most languages, it is possible to improve the accuracy of correcting non-word spelling mistakes in other languages by using the equation of this article.

The limitations of our study is that, in the Persian language, there is no any spelling correction system in the field of medicine, so we were unable to properly compare our results with previous results. Another limitation of our system is that it focuses only on detecting and correcting non-word spelling errors. In other words, the presented method in this paper is not capable of detecting and correcting real-word spelling mistakes.

Conclusion

The results indicated that high-quality spelling correction is possible on clinical reports. The system also achieved significant savings during the length of the writing process and final approval of reports in the imaging department. Since we focused on non-word errors in this article, future works may also include the study of real-word errors and developing misspelling detection and correction system for other kinds of medical texts.

Funding information This study was funded by the Cancer Research Center of Cancer Institute of Iran/Sohrabi cancer charity (grant number 96-34375-51-01).

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Ethical Approval For this type of study, formal consent is not required.

References

1. Holzinger, A., et al., *Biomedical text mining: state-of-the-art, open problems and future challenges,* in *Interactive knowledge discovery and data mining in biomedical informatics*. 2014, Springer. p. 271–300.
2. Wong W, Glance D: Statistical semantic and clinician confidence analysis for correcting abbreviations and spelling errors in clinical progress notes. Artificial intelligence in medicine 53(3):171–180, 2011
3. Zhou L et al.: Analysis of errors in dictated clinical documents assisted by speech recognition software and professional transcriptionists. JAMA Network Open 1(3):e180530–e180530, 2018
4. Turchin, A., et al. *Identification of misspelled words without a comprehensive dictionary using prevalence analysis*. in *AMIA Annual Symposium Proceedings*. 2007. American Medical Informatics Association.
5. Dalianis, H., *Clinical Text Mining: Secondary Use of Electronic Patient Records*. 2018: Springer.
6. Dalianis, H., *Clinical text retrieval-an overview of basic building blocks and applications*, in *Professional Search in the Modern World*. 2014, Springer. p. 147–165.
7. Ringler MD, Goss BC, Bartholmai BJ: Syntactic and semantic errors in radiology reports associated with speech recognition software. Health informatics journal 23(1):3–13, 2017
8. Zech, J., et al., Detecting insertion, substitution, and deletion errors in radiology reports using neural sequence-to-sequence models. Annals of Translational Medicine, 2018.
9. Zhang, Y. *Contextualizing consumer health information searching: an analysis of questions in a social Q&A community*. in *Proceedings of the 1st ACM International Health Informatics Symposium*. 2010. ACM.
10. Golkar, A., et al. *Improve word sense disambiguation by proposing a pruning method for optimizing conceptual density's contexts*. in *Artificial Intelligence and Signal Processing (AISP), 2015 International Symposium on*. 2015. IEEE.
11. Sarker, A. and G. Gonzalez-Hernandez, *An unsupervised and customizable misspelling generator for mining noisy health-related text sources.* arXiv preprint arXiv:1806.00910, 2018.
12. Nizamuddin, U. and H. Dalianis. *Detection of spelling errors in Swedish clinical text*. in *1st Nordic workshop on evaluation of spellchecking and proofing tools (NorWEST2014), SLTC* 2014, Uppsala. 2014.
13. Dalianis, H., *Characteristics of Patient Records and Clinical Corpora*, in *Clinical Text Mining*. 2018, Springer. p. 21–34.
14. Hussain, F. and U. Qamar. *Identification and Correction of Misspelled Drugs Names in Electronic Medical Records (EMR)*. in *ICEIS (2)*. 2016.
15. Kilicoglu H et al.: *An ensemble method for spelling correction in consumer health questions*. in *AMIA Annual Symposium Proceedings*. 2015. American Medical Informatics Association.

16. Zhou, X., et al., *Context-sensitive spelling correction of consumer-generated content on health care.* JMIR medical informatics, 2015. 3(3).
17. Ruch P, Baud R, Geissbühler A: Using lexical disambiguation and named-entity recognition to improve spelling correction in the electronic patient record. Artificial intelligence in medicine 29(1–2): 169–184, 2003
18. Siklósi, B., A. Novák, and G. Prószéky. *Context-aware correction of spelling errors in Hungarian medical documents*. in *International Conference on Statistical Language and Speech Processing*. 2013. Springer.
19. Grigonyté, G., et al. *Improving readability of Swedish electronic health records through lexical simplification: First results*. in *European Chapter of ACL (EACL), 26-30 April, 2014, Gothenburg, Sweden*. 2014. Association for Computational Linguistics.
20. Tolentino HD et al.: A UMLS-based spell checker for natural language processing in vaccine safety. BMC medical informatics and decision making 7(1):3, 2007
21. Doan S, Bastarache L, Klimkowski S, Denny JC, Xu H: Integrating existing natural language processing tools for medication extraction from discharge summaries. Journal of the American Medical Informatics Association 17(5):528–531, 2010
22. Lai KH, Topaz M, Goss FR, Zhou L: Automated misspelling detection and correction in clinical free-text records. Journal of biomedical informatics 55:188–195, 2015
23. Fivez, P., S. Šuster, and W. Daelemans, *Unsupervised Context-Sensitive Spelling Correction of English and Dutch Clinical Free-Text with Word and Character N-Gram Embeddings.* arXiv preprint arXiv:1710.07045, 2017.
24. Pérez A, Atutxa A, Casillas A, Gojenola K, Sellart Á: Inferred joint multigram models for medical term normalization according to ICD. International journal of medical informatics 110:111–117, 2018
25. D'hondt, E., C. Grouin, and B. Grau. *Low-resource OCR error detection and correction in French Clinical Texts*. in *Proceedings of the Seventh International Workshop on Health Text Mining and Information Analysis*. 2016.
26. Faili H et al.: Vafa spell-checker for detecting spelling, grammatical, and real-word errors of Persian language. Literary and Linguistic Computing 31(1):95–117, 2014
27. Dowsett, D., *Radiological sciences dictionary: keywords, names and definitions*. 2009: CRC Press.
28. Damerau FJ: A technique for computer detection and correction of spelling errors. Communications of the ACM 7(3):171–176, 1964
29. Yazdani A, Safdari R, Golkar A, R Niakan Kalhori S: Words prediction based on N-gram model for free-text entry in electronic health records. Health information science and systems 7(1):6, 2019
30. Brown PF et al.: Class-based n-gram models of natural language. Computational linguistics 18(4):467–479, 1992
31. Minn MJ, Zandieh AR, Filice RW: Improving radiology report quality by rapidly notifying radiologist of report errors. Journal of digital imaging 28(4):492–498, 2015
32. Dashti SM: Real-word error correction with trigrams: correcting multiple errors in a sentence. Language Resources and Evaluation. 52(2):485–502, 2018 Jun 1
33. Kruskal JB, Reedy A, Pascal L, Rosen MP, Boiselle PM: Quality initiatives: lean approach to improving performance and efficiency in a radiology department. Radiographics. 32(2):573–587, 2012 Mar 5