

DOI: <https://doi.org/10.17816/DD321420>

# Технология распознавания речи в лучевой диагностике

Н.Д. Кудрявцев<sup>1</sup>, К.А. Бардасова<sup>2</sup>, А.Н. Хоружая<sup>1</sup><sup>1</sup> Научно-практический клинический центр диагностики и телемедицинских технологий, Москва, Российская Федерация<sup>2</sup> Уральский государственный медицинский университет, Екатеринбург, Российская Федерация

## АННОТАЦИЯ

Устройства, способные распознавать речь, являются перспективным инструментом для системы здравоохранения. Технология распознавания речи имеет довольно длинную историю применения в западных системах здравоохранения (с 1970-х годов), однако широкое распространение она получила лишь в начале XXI века, заменив медицинских транскрипционистов. Для отечественного здравоохранения данная технология относительно новая. Её активная разработка началась лишь в начале 2010-х годов, а повсеместное внедрение в здравоохранение — в конце 2010-х годов. Такая задержка связана с особенностями русского языка и ограничением вычислительных мощностей, присутствующих в начале XXI века.

В настоящее время комплексы устройств и программного обеспечения для распознавания речи используются в голосовом заполнении медицинской документации и позволяют сократить время подготовки протоколов рентгенологических исследований при сравнении с традиционным (клавиатурным) вводом текста.

В литературном обзоре отражена краткая история развития и применения технологии распознавания речи в лучевой диагностике. Отражены ключевые научные исследования, подтверждающие эффективность её использования в западных системах здравоохранения. Продемонстрирован отечественный опыт применения технологии распознавания речи и оценена её эффективность. Описаны перспективы дальнейшего развития данной технологии в российском здравоохранении.

**Ключевые слова:** научный обзор; технология распознавания речи; голосовой ввод; лучевая диагностика; рентгенология; заполнение медицинской документации.

## Как цитировать

Кудрявцев Н.Д., Бардасова К.А., Хоружая А.Н. Технология распознавания речи в лучевой диагностике // *Digital Diagnostics*. 2023. Т. 4, № 2. С. 185–196.DOI: <https://doi.org/10.17816/DD321420>

DOI: <https://doi.org/10.17816/DD321420>

# Speech recognition technology in radiology

Nikita D. Kudryavtsev<sup>1</sup>, Kristina A. Bardasova<sup>2</sup>, Anna N. Khoruzhaya<sup>1</sup>

<sup>1</sup> Moscow Center for Diagnostics and Telemedicine, Moscow, Russian Federation

<sup>2</sup> Ural State Medical University, Ekaterinburg, Russian Federation

## ABSTRACT

Speech recognition devices are promising tools for the healthcare system. Speech recognition technology has had a relatively long history of use in Western healthcare systems since the 1970s. However, it became widely used at the beginning of the 21<sup>st</sup> century, replacing medical transcriptionists. This technology is relatively new in home healthcare. Its active development began only in the early 2010s, and its implementation in healthcare started in late 2010. This delay is due to the idiosyncrasies of the Russian language and the limited computational power present at the beginning of the 21<sup>st</sup> century.

Currently, complexes of devices and software for speech recognition are used in the voice filling of medical records and can reduce the time for preparing reports for radiological examinations compared with traditional (keyboard) text input.

The literature review provides a brief history of speech recognition technology development and application in radiography. Key scientific studies showing its efficacy in Western healthcare systems are reflected. Voice recognition technology in the home is demonstrated, and its effectiveness is evaluated. The prospects for further development of this technology in Russian healthcare are described.

**Keywords:** medical records; radiation diagnostics; radiology; speech recognition software; voice input.

## To cite this article

Kudryavtsev ND, Bardasova KA, Khoruzhaya AN. Speech recognition technology in radiology. *Digital Diagnostics*. 2023;4(2):185–196.

DOI: <https://doi.org/10.17816/DD321420>

Received: 17.03.2023

Accepted: 19.04.2023

Published: 25.05.2023

DOI: <https://doi.org/10.17816/DD321420>

# 语音识别技术在放射诊断中的应用

Nikita D. Kudryavtsev<sup>1</sup>, Kristina A. Bardasova<sup>2</sup>, Anna N. Khoruzhaya<sup>1</sup>

<sup>1</sup> Moscow Center for Diagnostics and Telemedicine, Moscow, Russian Federation

<sup>2</sup> Ural State Medical University, Ekaterinburg, Russian Federation

## 简评

能够进行语音识别的设备是保健系统的一个有前途的工具。语音识别技术在西方医疗系统中有相当长的使用历史（自20世纪70年代以来），但它在21世纪初才得到了广泛推广，取代了医疗抄写员。对于国内的医疗保健来说，该技术是相对较新的。它的积极开发是在2010年代初才开始，并2010年代末才在保健事业广泛采用的。这种延迟是由于俄语的特点和21世纪初计算能力的限制而导致的。

语音识别的设备和软件包现在被用于通过语言输入填写病历，此外，与传统（用键盘）文本输入相比，减少了准备X射线学协议所需的时间。

本文综述简要介绍了语音识别技术在放射诊断中的发展和应用的历史。介绍了证实其在西方医疗系统中使用的有效性的主要科学研究。展示了国内使用语音识别技术的经验，并对其有效性进行了评估。描述了该技术在俄罗斯保健事业进一步发展的前景。

**关键词：**科学综述；语音识别技术；语音输入；放射诊断；X射线学；填写病历。

## To cite this article

Kudryavtsev ND, Bardasova KA, Khoruzhaya AN. 语音识别技术在放射诊断中的应用. *Digital Diagnostics*. 2023;4(2):185–196.

DOI: <https://doi.org/10.17816/DD321420>

收到: 17.03.2023

接受: 19.04.2023

发布日期: 25.05.2023

## INTRODUCTION

At present, voice control has become a standard feature for many home smart devices. This was made possible by the development of speech recognition technology (SRT), which can be used in systems analyzing and transforming the user's speech into digital data. Furthermore, to controlling smart devices, SRT has become very popular in telephone communications. Currently, when calling many government and commercial organizations, the user is met by an automatic responding machine that recognizes the caller's voice request and refers them to a selected specialist. In 2019, in Moscow, a project was launched to notify citizens of an appointment with a doctor and remind them of regular check-up using a voice assistant. During a call, the citizen could make an appointment with a medical specialist, cancel, or reschedule the visit, and the system also asked the patient about complaints.<sup>1</sup>

In healthcare, SRT systems are actively used in the voice filling of medical records. This is because healthcare professionals spend most of their working time preparing medical documentation [1–4]. This factor negatively affects the quality of medical care, particularly considering the limited duration of appointments for each patient. Opportunities for using this technology, for example, in diagnostic radiology, are associated with decreased time spent filling out protocols for diagnostic examinations and increased time analyzing diagnostic images and medical records and communicating with patients. Thus, voice input systems have become the most popular tool in imaging departments because their workflows are the most convenient for the implementation of such technology. Current systematic reviews [5–7] have shown that SRT systems are effective in these conditions, and good implementability is explained by large volumes of textual information that radiologists are required to record in protocols.

## EVOLUTION OF SPEECH RECOGNITION TECHNOLOGY IN RADIOLOGY

### Early years

Medical use of SRTs was first attempted in the 1970s and 1980s. In 1975, a prototype SRT system was described. It could analyze an extremely limited amount of medical vocabulary and terminology [8]. The technology was first tested in diagnostic radiology in 1981 [9]. Beth Israel Hospital started to use it for preparing protocols for X-ray examinations. Initially, local specialists used the coded language information processing (CLIP) system, which is a hierarchical standardized language of medical terms developed by M. Simon and BW Leeming [10]. This language

contained specially encoded medical terms. For example, value A was used for the anatomical region, A6 for the lower limb, and A61 for the thigh. Values B, B6, B61, B611, and B612 encoded bones, bones of the lower limbs, femur, lesser trochanter, and greater trochanter, respectively. By keyboarding various code combinations, a radiologist can prepare a protocol of any complexity and volume [11].

A main disadvantage of those SRT systems is a limited vocabulary (approximately 200 unique medical terms) because of the small random-access memory (RAM). This factor did not allow making a full description of diagnostic examinations. The combined use of STLs with the CLIP system was one of the solutions for the limited RAM.

In 1981, keyboard and voice inputs were compared [9]. The speed and quality of protocol preparation were evaluated for 60 imaging examinations. Consequently, the length of the protocol did not depend on the filling method, which indirectly confirmed the similar complexity of the examinations described. The period of filling out the protocol using voice input was four times longer than the period of the keyboard input. Voice-filled protocols contained an average of 12 recognition errors; whereas protocols prepared using the keyboard had none. This study also described some other limitations of the technology. The first systems could not completely suppress external noise; thus, the quality of medical speech recognition was low, which contributed to errors. The increased period of filling out a medical document was related to the system's ability to recognize only separately spoken words. A specialist had to pause between words and between codes if the CLIP system was used. This voice-filling method was uncomfortable and unnatural for human communication. Moreover, the voice input system required 3–6 h of preconfiguration and adaptation to the specific speech of the user (specialist).

All these limitations prevented those SRT systems from wider use in healthcare. Thus, medical transcriptionists were highly sought at that time. Nevertheless, all these attempts formed a basis for considering "weak points" of the technology at subsequent stages of development.

The increasing availability of imaging examinations, emergence of computed and magnetic resonance tomographs, and transition from analog to digital media have increased the workload of radiologists and duration of protocol preparation. In the mid-80s, in Western countries, audio transcription centers were opened to solve this problem and optimize human resources. A radiologist taped descriptions of findings identified during the interpretation of diagnostic images using audio recorders. Audio recordings were transferred to the audio transcription center, where medical transcriptionists transcribed the audio recordings and prepared the research protocols in text. The prepared protocols were checked and endorsed by the radiologist.

<sup>1</sup> Official website of the Moscow Mayor [Web]. The voice assistant will collect patient complaints by phone before a clinic appointment. Available from: <https://www.mos.ru/news/item/89302073/>.

In some cases, they were returned to the transcriptionists for error correction [12]. Dictaphone centers were widely used in foreign healthcare institutions until the 2010s [5, 13].

Some authors compared the efficiency of preparing imaging protocols using medical transcriptionists and SRT. They concluded that the main advantage of a medical transcriptionist is the ability to notice grammatical errors and consider contextual information about the patient. These advantages allow them to correctly understand and adjust audio recordings, even if the quality is poor [14]. However, given the scarcity of medical transcriptionists, printed protocols were often returned to the radiologist only 16 h after dictation [15]. Depending on the capacities of audio transcription centers, transcription of audio recordings took 6–24 h.

Later, medical transcriptionists were no longer needed because of further developments in SRT systems [14]. Despite the lower recognition accuracy of SRTs than medical transcriptionists, SRTs can reduce the monthly economic costs of the radiology department by 81% and significantly reduce the time for preparing examination protocols [5, 13, 16].

In the Russian healthcare system, attempts have also been made to introduce voice recorder centers,<sup>2</sup>; however, no open-access studies have evaluated their effectiveness. Owing to the lack of such centers in the modern Russian healthcare system, their use was considered inappropriate.

### 1990s: A new round of innovation

At the end of the 20<sup>th</sup> century, the amount of memory and vocabulary in SRT systems had increased (up to 19,000 radiology terms). The pre-setting time was reduced to several minutes, and the recognition accuracy was increased. In 1995, the first natural language transcription programs were introduced in the USA, allowing the detection of continuous English speech. Currently, radiologists could dictate at a comfortable conversational pace, without pausing between words. At this stage, these programs were less accurate than systems with separate input [17]. Further technological advances and increased speech recognition accuracy allowed the creation of continuous voice input systems instead of separate input systems.

### 21<sup>ST</sup> Century

In Western countries, automated SRTs started to be widely introduced and used in radiology departments in the early

2000s. Foreign studies have compared the speed of voice and keyboard filling of English-language medical records, volume of a medical document prepared, and satisfaction level of specialists. Data showed that technology leads to an increase in document filling speed by 26% and an increase in the volume of protocols. Voice input also allowed optimizing the workflow by reducing the time preparing medical records and improving the quality of protocol content. Specialists showed increased satisfaction from working with such documents [16, 18]. The frequency of errors also decreased, and most errors were related to punctuation [7]. The above factors led to a decrease in the total time for protocol preparation from 16 to 5 h [15]. Studies evaluating the time spent on protocol preparation have revealed that the average number of characters per minute and number and frequency of errors reflect the positive evolution of SRTs and their wider implementation in radiology departments. The percentage of protocols completed within 1 h increased from 26% to 58%. Protocol content became more structured [19].

Costs have also decreased over time. Thus, 5 of 7 cost-evaluating systematic reviews [5] have reported a decrease in costs, and two have reported an increase in costs. Similar results have been obtained in endocrinology and psychiatry, where SRT systems have improved the productivity and efficiency of HCPs [20]. The use of these systems in surgery helped reduce the time for protocol preparation from 4 to 3 days. The number of protocols prepared within 1 day increased from 22% to 37% [21]. From 2019 to the 2020s, foreign studies have concluded that SRTs save HCP time, increases HCP efficiency, and allows them to document more important details when filling out medical papers [22–24]. However, the main barrier to voice input system implementation can be considered a human factor, which is related to HCP resistance to change and fear of new technologies [25].

In 2016, a Microsoft research study showed that SRT systems have an accuracy of 94% and corresponded to the human level [26]. At present, this technology is widely used in medical practice in English-speaking countries, and the implementation level of voice input in radiology departments has reached 85%.<sup>3</sup> Currently, the market share of such programs in healthcare is approximately 25% globally.<sup>4</sup> Nuance Communications, IBM, and Philips are leaders in speech recognition software development.<sup>5</sup>

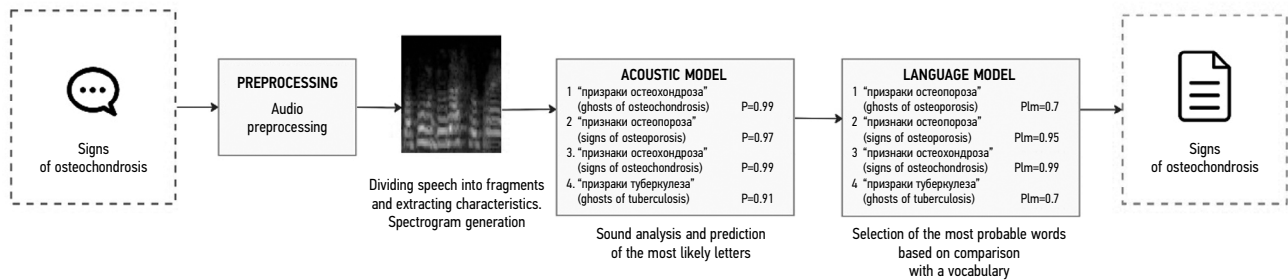
In approximately 45 years, speech recognition has become widespread in healthcare in English-speaking countries.

<sup>2</sup> Official site of State Budgetary Healthcare Institution "Infectious Clinical Hospital No. 1 of Moscow Health Department" [Internet]. History of the Hospital, Available from: <https://ikb1.ru/about/>.

<sup>3</sup> Reaction Data [Web]. Speech Rec in Radiology-State of the Market. 2019 [cited 2019 Dec 23]. Available from: <https://www.reactiondata.com/report/speech-recognition-in-radiology-state-of-the-market/>.

<sup>4</sup> Grand View Research [Web]. Voice And Speech Recognition Market Size Report, 2030. Available from: <https://www.grandviewresearch.com/industry-analysis/voice-recognition-market>.

<sup>5</sup> Nuance Communications. Healthcare Clinical Documentation AI Solutions & Services for the NHS (<https://www.nuance.com/en-gb/healthcare.html>); Philips. Healthcare--Philips (<https://www.dictation.philips.com/gb/industries/industry/healthcare-professionals/>); IBM. Watson Speech to Text (<https://www.ibm.com/cloud/watson-speech-to-text>).



**Fig. 1.** A simplified scheme of the operation of a classical speech recognition system. An algorithm for recognizing the “signs of osteochondrosis” phrase is presented.

It covered all healthcare levels, from primary care and emergency care to tertiary care clinics. State-of-art medical SRT systems for English-speaking users have an accuracy of up to 99%, can be adapted to different accents, and do not require machine training with the specialist’s voice profile.<sup>6</sup>

## HOW DOES A STATE-OF-THE-ART SPEECH RECOGNITION SYSTEM WORK?

As mentioned earlier, SRT involves translating human speech into text using a computer. Modern voice input systems use artificial intelligence (AI) algorithms that can significantly improve the quality and speed of user speech recognition [27]. The recognition process consists of several stages, with their characteristics (Fig. 1):

1. *Receiving an audio signal.* Sound recording is the first stage of speech recognition. It can be performed using the built-in microphone in any audio recording device, such as a smartphone. Then, the sound wave is converted into a digital format so that it can be processed by a computer.
2. *Audio preprocessing.* The received audio data are preprocessed to eliminate external background noise and make the user’s speech clearer. This improves the quality of speech recognition.
3. *Splitting into fragments.* The audio file is then split into small fragments of 10–25 ms. This is necessary to optimize audio data analysis. Each fragment is analyzed by the SRT system separately.
4. *Extraction of speech characteristics.* At this stage, speech parameters are calculated, including frequency, duration, and amplitude of sounds. They are used to identify phonemes that make up spoken words.
5. *Comparison with recognition models.* Phonemes are matched using acoustic models that are trained on many speech samples. These models can use various techniques including Hidden Markov models, neural

networks, and other machine-learning algorithms [28–30].

6. *Composition of words and text.* The language model combines phonemes into words and phrases and then into full text. This text may require additional processing to correct spelling errors and other inaccuracies.
7. *Inserting the generated text.* This is the final stage. The generated text is inserted into the medical document. Considering that speech recognition systems have not yet reached 100% accuracy, in some cases, the radiologist should manually adjust the recognized text.

To enable the system to recognize sounds regardless of sex, age, and intonation of the dictator and convert them into letters with greater accuracy, acoustic and language models use AI modules. Developers train neural networks on a dataset. The dataset includes various audio recordings and text examples. After receiving a voice signal, the neural network searches for a match in the database. The neural network continues its learning during the use and creates new combinations of “sound–letter” pairs, which makes it more likely to reproduce the intended text without spelling errors. During the learning process, the computer recognizes the most important features of the pronunciation of phonemes and records the received data as a user profile [31].<sup>7</sup>

Recently, the so-called end-to-end approaches to building SRT systems have become widespread. End-to-end approaches in SRT are methods of automatic processing of all speech signals, without performing separate processing steps such as feature extraction and speech recognition model generation [32]. Deep learning is one of the most popular end-to-end approaches. In this case, the neural network is trained directly on raw audio data without preprocessing [32]. Recurrent neural networks or convolutional neural networks are other examples of end-to-end approaches for speech recognition. In this case, the neural network is trained on the input signal and output text using the supervised learning technique [32].

<sup>6</sup> Nuance Communications [Internet]. Dragon Medical One--#1 Clinical Documentation Companion. Available from: <https://www.nuance.com/health-care/provider-solutions/speech-recognition/dragon-medical-one.html>.

<sup>7</sup> Cloud.mts.ru [Internet]. Speech recognition technology and its role for business. Available from: <https://cloud.mts.ru/cloud-thinking/blog/tekhnologiya-raspoznavaniya-rechi/>.



End-to-end approaches can provide higher recognition accuracy because the neural network is trained using all speech information. However, these approaches can be more difficult to implement and require more training data [33].

## RUSSIAN EXPERIENCE IN THE USE OF SPEECH RECOGNITION TECHNOLOGY IN RADIOLOGY

The first speech recognition systems for the Russian language appeared in the mid-2000s [34, 35]; however, the use of a general colloquial vocabulary did not allow the use of such systems in medical practice. The development of voice input systems that allows transcribing Russian speech with medical terms took several years [36]. The active development of SRTs was started only in the later 2010s. Such a delay was associated with difficulties in recognizing Russian speech.

The Russian language has a more complex structure of word formation than English because it is a synthetic language with many word forms. To recognize words, a larger vocabulary must be used; however, this slows down system performance [37]. For example, modern SRT systems for English-speaking users use a dictionary containing up to 300,000 words and terms, and for the Russian language, the vocabulary can contain more than 5 million words, word forms, and phrases [33, 38, 39]. In addition, most forms of the same word differ only in endings, which are often vaguely pronounced by users. This leads to an error in recognizing the entire phrase and the need to correct the final document. The Russian language has more options for sentence arrangement, whereas the English language uses strict grammatical constructions. This makes it difficult to create language models of the SRT system and reduces work accuracy.

In Russia, the Speech Technology Center Group (STC) is the leading developer of voice input systems for healthcare.<sup>8</sup> The first study of the effectiveness of SRTs in radiology departments was conducted in 2020 in seven city clinics of the Moscow Department of Health using the early version of Voice2Med voice input system (STC Group) with a recognition accuracy of 93%. The study compared the speed of completing medical records using keyboard input and an SRT system. Radiologists filled out protocols of computed and magnetic resonance imaging examinations. The time study showed that the average period of describing one examination using keyboard input was 10 min 15 s, and for an SRT system, it was 8 min 2 s. At the current stage of development, the accuracy of recognition of medical terms in Russian has reached 98%. This became possible thanks to preparing a

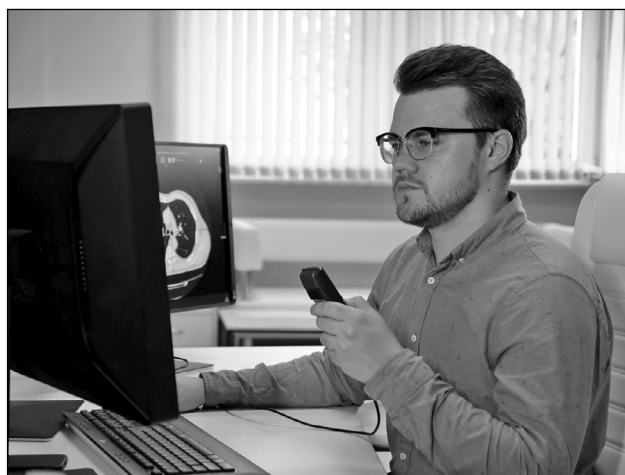


Fig. 2. Workplace of a radiologist at the Moscow Reference Center for Radiation Diagnostics, equipped with a speech recognition system. The process of filling medical records.

vocabulary of medical terms, based on 2.5 million imaging protocols, and analyzing feedback from radiologists<sup>9</sup> (Fig. 2).

In 2022, a survey of radiologists showed that 62.8% of the respondents noted an increase in their efficiency when using an SRT system. Most specialists who use voice input routinely rate the quality of recognition of radiological terms as good or excellent. Respondents note that when extraneous speech was recognized, the endings of words were recognized incorrectly. Moreover, the quality of recognition can be negatively affected by external background noise (working diagnostic equipment and communication of medical personnel with a patient or with colleagues) and low-quality sound recording devices. Important factors in new technology loyalty were the age of HCPs and their interest in innovation. Young professionals are more open to technology, and professionals aged 30–40 years are also more likely to use voice input systems in their work. Surveys showed a positive trend in the attitude of radiologists to SRTs within 2 years from the launch [40].

## OPPORTUNITIES FOR THE DEVELOPMENT OF SPEECH RECOGNITION TECHNOLOGY

Improvements in recognition accuracy will further reduce the preparation time of electronic medical records. One of the main tasks facing SRT system developers is to ensure high accuracy of speech analysis in difficult acoustic conditions, when the recording contains numerous noise or voices. As already mentioned, owing to some special characteristics of the Russian language, recognizing word endings become one of the most difficult tasks. Therefore, for Russian SRT

<sup>8</sup> Speech Technology Center Group [Internet]. Speech synthesis and recognition, recording and analysis, face and voice identification. Available from: <http://www.speechpro.ru/>.

<sup>9</sup> Speech Technology Center Group [Internet]. Voice2Med: Program for voice filling of medical records. Available from: <https://www.speechpro.ru/product/programmy-dlya-raspoznaniya-rechi-v-tekst/voice2med>.

systems, a language model that can predict and match words in sentences with high accuracy is needed.

The integration of voice input programs with medical information systems will allow remote filling of structured electronic medical records. As the system is developing, it cannot only recognize HCP phrases but also understand in which section of the medical record the recognized text should be placed. These functions will allow ultrasound diagnostic specialists, pathologists, endoscopists, and surgeons to fill out medical records directly during the medical intervention, not later, and this will significantly affect the quality of documents and the speed of their preparation.

SRT also has great potential in standardizing and unifying the vocabulary used in preparing medical records, including radiological protocols. Currently, no single list of terms has described the same pathological condition in radiology [41]. Even two radiologists working in the same department may use different synonyms to describe the same finding when preparing protocols. Some papers noted that the use of structured and standardized protocols with unified terms simplifies the perception of the necessary information by both other radiologists and other specialists [42–44].

To date, some researchers have attempted to develop a thesaurus to standardize the description of abnormal changes detected in computed tomography. A thesaurus contains 120 Russian radiological terms and examples of their description [45]. However, the development of a thesaurus is a complex task requiring consistent terminology agreed by numerous specialists and the radiology community.

## CONCLUSION

The literature review provides a brief historical background on the development of SRTs in radiology departments, describes their evolution in detail, and evaluates the advantages and disadvantages of SRT based on literature data. Special attention is paid to the use of SRTs in Russian radiology departments. Some studies have demonstrated a significant improvement in the accuracy of recognition of Russian medical terms.

## REFERENCES

1. Vechorko VI. Distribution of working time at an outpatient appointment of a district therapist with a nurse in a polyclinic in Moscow (photochronometric observation). *Social Aspects Public Health*. 2016;(6):4. (In Russ).
2. Kaplieva OV, Marega LA, Vorobyeva LP. Timekeeping of working hours of doctors of the children's consultative and diagnostic department. *Far Eastern Med J*. 2018;(4):72–76. (In Russ).
3. Ryabchikov IV, Zagafarov RR, Mukhina VV, et al. Distribution of the traumatologist-orthopaedician's working time with outpatients. *Moscow Sur J*. 2018;(6):38–43. (In Russ). doi: 10.17238/issn2072-3180.2018.6.38-43
4. Kudryavtsev ND, Sergunova KA, Ivanova GV, et al. Evaluation of the effectiveness of the implementation of speech recognition technology for the preparation of

In the future, the use of such technologies can reduce the preparation period of medical records to allow spending more time communicating with patients and analyzing their medical histories. This opens new opportunities for personalized healthcare development. However, errors remain in ending recognition and word agreement in a sentence, and specialists must spend more time correcting them. In the future, new AI algorithms can solve these problems.

Some studies have demonstrated the positive attitude of radiologists to SRT systems, which is manifested in their more frequent use in their work. This technology should be further developed in the Russian healthcare sector because pilot national and well-established foreign projects indicate positive changes. Further improving the accuracy of medical term recognition will attract even more SRT supporters among HCPs.

## ADDITIONAL INFORMATION

**Funding source.** This article was prepared by a group of authors as a part of the research and development effort titled "Theoretical and methodological framework for digital transformation in radiology" (USIS No. 123031400118-0) in accordance with the Order No. 1196 dated December 21, 2022 "On approval of state assignments funded by means of allocations from the budget of the city of Moscow to the state budgetary (autonomous) institutions subordinate to the Moscow Health Care Department, for 2023 and the planned period of 2024 and 2025" issued by the Moscow Health Care Department.

**Competing interests.** The authors declare that they have no competing interests.

**Authors' contribution.** All authors made a substantial contribution to the conception of the work, acquisition, analysis, interpretation of data for the work, drafting and revising the work, final approval of the version to be published and agree to be accountable for all aspects of the work. N. Kudryavtsev — concept and design of the paper, text writing, approval of the final version of the paper; K. Bardasova — data collection and analysis; A. Khoruzhaya — data interpretation, text editing.

radiological protocols. *VIT*. 2020;6(S1):58–64. (In Russ). doi: 10.37690/1811-0193-2020-S1-58-64

5. Blackley SV, Huynh J, Wang L, et al. Speech recognition for clinical documentation from 1990 to 2018: A systematic review. *J Am Med Inf Association*. 2019;26(4):324–338. doi: 10.1093/jamia/ocy179

6. Motyer RE, Liddy S, Torreggiani WC, Buckley O. Frequency and analysis of non-clinical errors made in radiology reports using the National Integrated Medical Imaging System voice recognition dictation software. *Ir J Med Sci*. 2016;185(4):921–927. doi: 10.1007/s11845-016-1507-6

7. Hodgson T, Coiera E. Risks and benefits of speech recognition for clinical documentation: A systematic review. *J Am Med Inf Association*. 2016;23(e1):e169–e179. doi: 10.1093/jamia/ocv152



8. Itakura F. Minimum prediction residual principle applied to speech recognition. *IEEE Trans Acoust Speech Signal Process.* 1975;23(1):67–72. doi: 10.1109/TASSP.1975.1162641
9. Leeming W, Porter D, Jackson JD, et al. Computerized radiologic reporting with voice data-entry. *Radiology.* 1981;138(3):585–588. doi: 10.1148/radiology.138.3.7465833
10. Simon M, Leeming BW, Bleich HL, et al. Computerized radiology reporting using coded language. *Radiology.* 1974;113(2):343–349. doi: 10.1148/113.2.343
11. Vogel M, Kaisers W, Wassmuth R, Mayatepek E. Analysis of documentation speed using web-based medical speech recognition technology: Randomized controlled trial. *J Med Internet Res.* 2015;17(11):e247. doi: 10.2196/jmir.5072
12. Ramaswamy MR, Chaljub G, Esch O, et al. Continuous speech recognition in MR imaging reporting. *Am J Roentgenol.* 2000;174(3):617–622. doi: 10.2214/ajr.174.3.1740617
13. Poder TG, Fisetto JF, Déry V. Speech recognition for medical dictation: Overview in quebec and systematic review. *J Med Systems.* 2018;42(5):89. doi: 10.1007/s10916-018-0947-0
14. Sankaranarayanan B, David G, Vishwanath KR, et al. Would technology obliterate medical transcription? In: Proceedings of the 2017 ACM SIGMIS Conference on Computers and People Research. New York, NY, USA: ACM; 2017. P. 97–104. doi: 10.1145/3084381.3084414
15. Houston JD, Rupp FW. Experience with implementation of a radiology speech recognition system. *J Digital Imaging.* 2000;13(3):124–128. doi: 10.1007/BF03168385
16. Saxena K, Diamond R, Conant RF, et al. Provider adoption of speech recognition and its impact on satisfaction, documentation quality, efficiency, and cost in an inpatient EHR. *AMIA Jt Summits Transl Sci Proc.* 2018;2017:186–195.
17. Schwartz LH, Kijewski P, Hertogen H, et al. Voice recognition in radiology reporting. *Am J Roentgenol.* 1997;169(1):27–29. doi: 10.2214/ajr.169.1.9207496
18. Vogel M, Kaisers W, Wassmuth R, Mayatepek E. Analysis of documentation speed using web-based medical speech recognition technology: Randomized controlled trial. *J Med Int Research.* 2015;17(11):e247. doi: 10.2196/jmir.5072
19. Hammana I, Lepanto L, Poder T, et al. Speech recognition in the radiology department: A systematic review. *Health Inf Manag.* 2015;44(2):4–10. doi: 10.1177/183335831504400201
20. Mohr DN, Turner DW, Pond GR, et al. speech recognition as a transcription aid: A randomized comparison with standard transcription. *J Am Med Inf Association.* 2003;10(1):85–93. doi: 10.1197/jamia.m1130
21. Singh M, Pal TR. Voice recognition technology implementation in surgical pathology: Advantages and limitations. *Arch Pathol Laboratory Med.* 2011;135(11):1476–1481. doi: 10.5858/arpa.2010-0714-OA
22. Goss FR, Blackley SV, Ortega CA, et al. A clinician survey of using speech recognition for clinical documentation in the electronic health record. *Int J Med Inf.* 2019;130:103938. doi: 10.1016/j.ijmedinf.2019.07.017
23. Blackley SV, Schubert VD, Goss FR, et al. Physician use of speech recognition versus typing in clinical documentation: A controlled observational study. *Int J Med Inform.* 2020;(141):104178. doi: 10.1016/j.ijmedinf.2020.104178
24. Yang L, Ene IC, Belaghi RA, et al. Stakeholders' perspectives on the future of artificial intelligence in radiology: A scoping review. *Eur Radiol.* 2022;32(3):1477–1495. doi: 10.1007/s00330-021-08214-z
25. European Society of Radiology (ESR). Impact of artificial intelligence on radiology: A EuroAIM survey among members of the European Society of Radiology. *Insights Imaging.* 2019;10(1):105. doi: 10.1186/s13244-019-0798-3
26. Szymański P, Żelasko P, Morzyet M, et al. WER we are and WER we think we are. *arXiv.* arXiv:2010.03432.2020. doi: 10.48550/arXiv.2010.03432
27. Li J. Recent advances in end-to-end automatic speech recognition. *arXiv.* arXiv:2111.01690.2022. doi: 10.48550/arXiv.2111.01690
28. Juang BH, Rabiner LR. Hidden markov models for speech recognition. *Technometrics.* 1991;33(3):251–272.
29. Graves A, Mohamed A, Hinton G. Speech recognition with deep recurrent neural networks In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. Vancouver, BC, Canada: IEEE; 2013. P. 6645–6649. doi: 10.48550/arXiv.1303.5778
30. Deng L, Li X. Machine learning paradigms for speech recognition: An overview. *IEEE Trans Audio Speech Lang Process.* 2013;21(5):1060–1089.
31. Kazachkin AE. Speech recognition methods, modern speech technologies. *Young Scientist.* 2019;(39):6–8. (In Russ).
32. Kamath U, Liu J, Whitaker J. Deep learning for NLP and speech recognition. Cham: Springer International Publishing; 2019. 621 p.
33. Wang D, Wang X, Lv S. An overview of end-to-end automatic speech recognition. *Symmetry.* 2019;11(8):1018. doi: 10.3390/sym11081018
34. Zhzhikashvili VA, Farkhadov MP, Petukhova NV, Zhzhikashvili AV. The first voice recognition applications in Russian language for use in The Interactive Information Systems. In: Speech and Computer. Saint-Petersburg, SPECOM; 2004. P. 304–307. (In Russ).
35. Karpov AA, Ronzhin AA, Li IV. SIRIUS system of dictoron-independent recognition of the merged Russian speech. *Izvestia Southern Federal University. Technical Sci.* 2005;54(10):44–54. (In Russ).
36. Irzaev MG. The use of voice input of information in medical institutions to fill in electronic charts and patient medical histories. *New technologies and techniques in medicine, biology and ecology: Collection of scientific papers.* 2013;(3):149–154. (In Russ).
37. Vazhenina D, Markov K, Karpov A, et al. State-of-the-art speech recognition technologies for Russian language. In: Proceedings of the 2012 Joint International Conference on Human-Centered Computer Environments. Aizu-Wakamatsu Japan: ACM; 2012. P. 59–63. doi: 10.1145/2160749.2160763
38. Kamvar M, Chelba C. Optimal size, freshness and time-frame for voice search vocabulary. *arXiv.* arXiv:1210.8436.2012.
39. Kipyatkova IS, Karpov AA. An analytical survey of large vocabulary Russian speech recognition systems. *SPIIRAS Proceedings.* 2014;1(12):7. (In Russ). doi: 10.15622/sp.12.1
40. Kudryavtsev ND, Semenov DS, Kozhikhina DD, Vladzmyrskyy AV. Speech recognition technology: Results of a survey of radiologists at the Moscow reference center for diagnostic radiology. *Healthcare Management.* 2022;8(3):95–104. (In Russ). doi: 10.33029/2411-8621-2022-8-3-95-104
41. Sinityn VE, Komarova MA, Mershina EA. Protocol of radiological description: Past, present, future. *Bulletin Radiology Radiology.* 2014;(3):35–40. (In Russ).
42. Sobez LM, Kim SH, Angstwurm M, et al. Creating high-quality radiology reports in foreign languages through multilingual structured reporting. *Eur Radiol.* 2019;29(11):6038–6048. doi: 10.1007/s00330-019-06206-8
43. Ganeshan D, Duong PA, Probyn L, et al. Structured reporting in radiology. *Academic Radiology.* 2018;25(1):66–73. doi: 10.1016/j.acra.2017.08.005
44. Dos Santos PD, Hempel JM, Mildemberger P, et al. Structured reporting in clinical routine. *Rofo.* 2019. Vol. 191, N 01. P. 33–39. doi: 10.1055/a-0636-3851
45. Andrianova MG, Kudryavtsev ND, Petryaykin AV. Development of a thesaurus of radiological terms for voice filling of diagnostic research protocols. *Digital Diagnostics.* 2022;3(S1):21–22. (In Russ). doi: 10.17816/DD105703

## СПИСОК ЛИТЕРАТУРЫ

1. Вечорко В.И. Распределение рабочего времени на амбулаторном приеме врача-терапевта участкового с медицинской сестрой в поликлинике города Москвы (фотохронометражное наблюдение) // Социальные аспекты здоровья населения. 2016. № 6. С. 4.
2. Каплиева О.В., Марега Л.А., Воробьева Л.П. Хронометраж рабочего времени врачей детского консультативно-диагностического отделения // Дальневосточный медицинский журнал. 2018. № 4. С. 72–76.
3. Ryabchikov I.V., Zagafarov R.R., Mukhina V.V., et al. Distribution of the traumatologist-orthopaedician's working time with outpatients // *Moscow Sur J.* 2018. N 6. P. 38–43. doi: 10.17238/issn2072-3180.2018.6.38-43
4. Kudryavtsev N.D., Sergunova K.A., Ivanova G.V., et al. Evaluation of the effectiveness of the implementation of speech recognition technology for the preparation of radiological protocols // *VIT.* 2020. Vol. 6, N S1. P. 58–64. doi: 10.37690/1811-0193-2020-S1-58-64
5. Blackley S.V., Huynh J., Wang L., et al. Speech recognition for clinical documentation from 1990 to 2018: A systematic review // *J Am Med Inf Association.* 2019. Vol. 26, N 4. P. 324–338. doi: 10.1093/jamia/ocy179
6. Motyer R.E., Liddy S., Torreggiani W.C., Buckley O. Frequency and analysis of non-clinical errors made in radiology reports using the National Integrated Medical Imaging System voice recognition dictation software // *Ir J Med Sci.* 2016. Vol. 185, N 4. P. 921–927. doi: 10.1007/s11845-016-1507-6
7. Hodgson T., Coiera E. Risks and benefits of speech recognition for clinical documentation: A systematic review // *J Am Med Inf Association.* 2016. Vol. 23, N e1. P. e169–e179. doi: 10.1093/jamia/ocv152
8. Itakura F. Minimum prediction residual principle applied to speech recognition // *IEEE Trans Acoust Speech Signal Process.* 1975. Vol. 23, N 1. P. 67–72. doi: 10.1109/TASSP.1975.1162641
9. Leeming W., Porter D., Jackson J.D., et al. Computerized radiologic reporting with voice data-entry // *Radiology.* 1981. Vol. 138, N 3. P. 585–588. doi: 10.1148/radiology.138.3.7465833
10. Simon M., Leeming B.W., Bleich H.L., et al. Computerized radiology reporting using coded language // *Radiology.* 1974. Vol. 113, N 2. P. 343–349. doi: 10.1148/113.2.343
11. Vogel M., Kaisers W., Wassmuth R., Mayatepek E. Analysis of documentation speed using web-based medical speech recognition technology: Randomized controlled trial // *J Med Internet Res.* 2015. Vol. 17, N 11. P. e247. doi: 10.2196/jmir.5072
12. Ramaswamy M.R., Chaljub G., Esch O., et al. Continuous speech recognition in MR imaging reporting // *Am J Roentgenol.* 2000. Vol. 174, N 3. P. 617–622. doi: 10.2214/ajr.174.3.1740617
13. Poder T.G., Fisette J.F., Déry V. Speech recognition for medical dictation: Overview in quebec and systematic review // *J Med Systems.* 2018. Vol. 42, N 5. P. 89. doi: 10.1007/s10916-018-0947-0
14. Sankaranarayanan B., David G., Vishwanath K.R., et al. Would technology obliterate medical transcription? // *Proceedings of the 2017 ACM SIGMIS Conference on Computers and People Research.* New York, NY, USA: ACM, 2017. P. 97–104. doi: 10.1145/3084381.3084414
15. Houston J.D., Rupp F.W. Experience with implementation of a radiology speech recognition system // *J Digital Imaging.* 2000. Vol. 13, N 3. P. 124–128. doi: 10.1007/BF03168385
16. Saxena K., Diamond R., Conant R.F., et al. Provider adoption of speech recognition and its impact on satisfaction, documentation quality, efficiency, and cost in an inpatient EHR // *AMIA Jt Summits Transl Sci Proc.* 2018. Vol. 2017. P. 186–195.
17. Schwartz L.H., Kijewski P., Hertogen H., et al. Voice recognition in radiology reporting // *Am J Roentgenol.* 1997. Vol. 169, N 1. P. 27–29. doi: 10.2214/ajr.169.1.9207496
18. Vogel M., Kaisers W., Wassmuth R., Mayatepek E. Analysis of documentation speed using web-based medical speech recognition technology: Randomized controlled trial // *J Med Int Research.* 2015. Vol. 17, N 11. P. e247. doi: 10.2196/jmir.5072
19. Hammana I., Lepanto L., Poder T., et al. Speech recognition in the radiology department: A systematic review // *Health Inf Manag.* 2015. Vol. 44, N 2. P. 4–10. doi: 10.1177/183335831504400201
20. Mohr D.N., Turner D.W., Pond G.R., et al. Speech recognition as a transcription aid: A randomized comparison with standard transcription // *J Am Med Inf Association.* 2003. Vol. 10, N 1. P. 85–93. doi: 10.1197/jamia.m1130
21. Singh M., Pal T.R. Voice recognition technology implementation in surgical pathology: Advantages and limitations // *Arch Pathol Laboratory Med.* 2011. Vol. 135, N 11. P. 1476–1481. doi: 10.5858/arpa.2010-0714-OA
22. Goss F.R., Blackley S.V., Ortega C.A., et al. A clinician survey of using speech recognition for clinical documentation in the electronic health record // *Int J Med Inf.* 2019. N 130. P. 103938. doi: 10.1016/j.ijmedinf.2019.07.017
23. Blackley S.V., Schubert V.D., Goss F.R., et al. Physician use of speech recognition versus typing in clinical documentation: A controlled observational study // *Int J Med Inform.* 2020. N 141. P. 104178. doi: 10.1016/j.ijmedinf.2020.104178
24. Yang L., Ene I.C., Belaghi R.A., et al. Stakeholders' perspectives on the future of artificial intelligence in radiology: A scoping review // *Eur Radiol.* 2022. Vol. 32, N 3. P. 1477–1495. doi: 10.1007/s00330-021-08214-z
25. European Society of Radiology (ESR). Impact of artificial intelligence on radiology: A EuroAIM survey among members of the European Society of Radiology // *Insights Imaging.* 2019. Vol. 10, N 1. P. 105. doi: 10.1186/s13244-019-0798-3
26. Szymański P., Żelasko P., Morzyet M., et al. WER we are and WER we think we are // *arXiv.* arXiv:2010.03432.2020. doi: 10.48550/arXiv.2010.03432
27. Li J. Recent advances in end-to-end automatic speech recognition // *arXiv.* arXiv:2111.01690.2022. doi: 10.48550/arXiv.2111.01690
28. Juang B.H., Rabiner L.R. Hidden markov models for speech recognition // *Technometrics.* 1991. Vol. 33, N 3. P. 251–272.
29. Graves A., Mohamed A., Hinton G. Speech recognition with deep recurrent neural networks // *2013 IEEE International Conference on Acoustics, Speech and Signal Processing.* Vancouver, BC, Canada: IEEE, 2013. P. 6645–6649. doi: 10.48550/arXiv.1303.5778
30. Deng L., Li X. Machine learning paradigms for speech recognition: An overview // *IEEE Trans Audio Speech Lang Process.* 2013. Vol. 21, N 5. P. 1060–1089.
31. Казачкин А.Е. Методы распознавания речи, современные речевые технологии // *Молодой ученый.* 2019. № 39. С. 6–8.
32. Kamath U., Liu J., Whitaker J. Deep learning for NLP and speech recognition. Cham: Springer International Publishing, 2019. 621 p.

33. Wang D., Wang X., Lv S. An overview of end-to-end automatic speech recognition // *Symmetry*. 2019. Vol. 11, N 8. P. 1018. doi: 10.3390/sym11081018
34. Zhzhikashvili V.A., Farkhadov M.P., Petukhova N.V., Zhzhikashvili A.V. The first voice recognition applications in Russian language for use in The Interactive Information Systems // *Speech and Computer*. Saint-Petersburg, SPECOM, 2004. P. 304–307.
35. Карпов А.А., Ронжин А.А., Ли И.В. SIRIUS система дикторне-зависимого распознавания слитной русской речи // *Известия Южного федерального университета. Технические науки*. 2005. Т. 54, № 10. С. 44–54.
36. Ирзаев М.Г. Использование голосового ввода информации в медицинских учреждениях для заполнения электронных карт и историй болезней пациентов // *Новые технологии и техника в медицине, биологии и экологии: сборник научных трудов*. 2013. № 3. С. 149–154.
37. Vazhenina D., Markov K., Karpov A., et al. State-of-the-art speech recognition technologies for Russian language // *Proceedings of the 2012 Joint International Conference on Human-Centered Computer Environments*. Aizu-Wakamatsu Japan: ACM, 2012. P. 59–63. doi: 10.1145/2160749.2160763
38. Kamvar M., Chelba C. Optimal size, freshness and time-frame for voice search vocabulary // *arXiv*. arXiv:1210.8436.2012.
39. Kipyatkova I.S., Karpov A.A. An analytical survey of large vocabulary Russian speech recognition systems // *SPIIRAS Proceedings*. 2014. Vol. 1, N 12. P. 7. doi: 10.15622/sp.12.1
40. Kudryavtsev N.D., Semenov D.S., Kozhikhina D.D., Vladzimirskyy A.V. Speech recognition technology: Results of a survey of radiologists at the Moscow reference center for diagnostic radiology // *Healthcare Management*. 2022. Vol. 8, N 3. P. 95–104. doi: 10.33029/2411-8621-2022-8-3-95-104
41. Синицын В.Е., Комарова М.А., Мершина Е.А. Протокол рентгенологического описания: прошлое, настоящее, будущее // *Вестник рентгенологии и радиологии*. 2014. № 3. С. 35–40.
42. Sobez L.M., Kim S.H., Angstwurm M., et al. Creating high-quality radiology reports in foreign languages through multilingual structured reporting // *Eur Radiol*. 2019. Vol. 29, N 11. P. 6038–6048. doi: 10.1007/s00330-019-06206-8
43. Ganeshan D., Duong P.A., Probyn L., et al. Structured reporting in radiology // *Academic Radiology*. 2018. Vol. 25, N 1. P. 66–73. doi: 10.1016/j.acra.2017.08.005
44. Dos Santos P.D., Hempel J.M., Mildemberger P., et al. Structured reporting in clinical routine // *Rofo*. 2019. Vol. 191, N 01. P. 33–39. doi: 10.1055/a-0636-3851
45. Андрианова М.Г., Кудрявцев Н.Д., Петряйкин А.В. Разработка тезауруса рентгенологических терминов для голосового заполнения протоколов диагностических исследований // *Digital Diagnostics*. 2022. Т. 3, № S1. С. 21–22. doi: 10.17816/DD105703

## AUTHORS' INFO

### \* Nikita D. Kudryavtsev;

address: 24/1 Petrovka street, 127051 Moscow, Russia;  
ORCID: 0000-0003-4203-0630;  
eLibrary SPIN: 1125-8637;  
e-mail: KudryavtsevND@zdrav.mos.ru

### Kristina A. Bardasova;

ORCID: 0009-0002-4310-1357;  
eLibrary SPIN: 1156-7627;  
e-mail: bardasovakris@mail.ru

### Anna N. Khoruzhaya;

ORCID: 0000-0003-4857-5404;  
eLibrary SPIN: 7948-6427;  
e-mail: KhoruzhayaAN@zdrav.mos.ru

## ОБ АВТОРАХ

### \* Кудрявцев Никита Дмитриевич;

адрес: Россия, 127051, Москва, ул. Петровка, д. 24, стр. 1;  
ORCID: 0000-0003-4203-0630;  
eLibrary SPIN: 1125-8637;  
e-mail: KudryavtsevND@zdrav.mos.ru

### Бардасова Кристина Алексеевна;

ORCID: 0009-0002-4310-1357;  
eLibrary SPIN: 1156-7627;  
e-mail: bardasovakris@mail.ru

### Хоружая Анна Николаевна;

ORCID: 0000-0003-4857-5404;  
eLibrary SPIN: 7948-6427;  
e-mail: KhoruzhayaAN@zdrav.mos.ru

\* Corresponding author / Автор, ответственный за переписку