# A Speech-to-Text Interface for MammoClass

Ricardo Sousa Rocha[*†], Pedro Ferreira[*], Inês Dutra[*†], Ricardo Correia[‡], Rogerio Salvini[‖], Elizabeth Burnside[¶]
[*]CRACS-INESC TEC, Porto, Portugal
[†]DCC-FC, Universidade do Porto, Porto, Portugal
[‡]CINTESIS and CIDES-FM, Universidade do Porto, Porto, Portugal
[¶]University of Wisconsin, Medical School, Madison, WI, USA
[‖]Instituto de Informática, Universidade Federal de Goiás, Goiás, Brazil

***Abstract*—Mammoclass is a web tool that allows users to enter a small set of variable values that describe a finding in a mammography, and produces a probability of this finding being malignant or benign. The tool requires that the user types in every variable a value in order to perform a prediction. In this work, we present a speech-to-text interface integrated to MammoClass that allows radiologists to speak up a mammography report instead of typing it in. This new MammoClass module can take audio content, transcribe it into written words, and automatically extract the variable values by applying a parser to the recognized text. Results of spoken mammography reports show that the same variables are extracted for both types of input: typed in or dictated text.**

***Index Terms*—speech recognition; parsing; machine learning; mammography; BI-RADS**

## I. Introduction

Speech recognition technology has been improved along the years, allowing for more accurate reporting, and consequent storage of more qualitative information. Several works in the literature report successful stories of using speech recognition to extract meaningful words from dictated texts. For example, Kang *et al.* [1] use speech recognition technology in surgical pathology and conclude that it is useful in their anatomic pathology workflow and provides a good return on investment, error reduction, and cost savings. Despite cases of success, speech recognition systems only work well when vocabularies are limited and dictation tasks are performed in isolated, dedicated workspaces, such as radiology or pathology [2]. They are much less suitable in noisy public spaces, where performance is poor and the confidentiality of patient health information is threatened [3].

Speech recognition is often confused with Voice recognition, which has other objectives. In fact, most references cited in this text use the term Voice recognition instead of Speech recognition. Voice recognition is usually related to recognize/identify an individual person/voice, and is independent of the language. Speech recognition is language dependent and strips out personal differences to detect words. Our work is in the context of Speech recognition. We are interested in recognizing and extracting relevant words from dictated texts in the area of radiology, more specifically, mammography terms. Some works in the literature do not favor the use of speech recognition technology in the area of radiology and report a high error rate on the resulting recognized texts (medical reports) [4], [5], [6], [7], [8]. A relatively recent work [9] tries reducing the error rate by successive revisions of dictated texts. These works focus on the text itself rather than on relevant words that could be extracted from the text to build structured data for posterior automatic studies. A new trend is to use digital pens to recognize annotations on special interactive paper [10]. This is orthogonal to our work and can be a complementary tool.

When describing findings in a mammography (breast X-Ray image), a radiologist uses a well-known terminology created by the American College of Radiology (ACR): the BI-RADS©(Breast Imaging and Reporting Data System) descriptors [11]. Based on these descriptors, on their own experience, and on other facts about the clinical patient record, physicians can then reach a conclusion about a diagnostic and usually produce a report. A BI-RADS category that ranks the degree of malignancy is also assigned to the finding. Structured BI-RADS descriptors are very useful since predictive models can be automatically built from data and from the physician's experience to support clinical decisions. From that point of view, although dictated reports have a high error rate regarding whole texts, they can help accelerating the process of building structured data, because the focus is on relevant words and not on the text as a whole. In the past, we have built a classifier to predict malignancy of a mammography finding, using BI-RADS descriptors (and other data) collected from the School of Medicine and Public Health of the University of Wisconsin, Madison, WI, USA. The training data consisted of a set of 348 patients whose BI-RADS descriptors and categories were annotated by experienced radiologists specialist in mammography [12]. Ground-truth (benign or malignant), taken from patient registries, was also known for those mammography findings. Our classifier, a Support Vector Machine (SVM), based on the SMO algorithm [13], trained with a Radial-Basis-Function (RBF) kernel, has a very good performance when compared with the radiologists'.

An interface to test and further assess the performance of the model, MammoClass, was built and is publicly available (http://cracs.fc.up.pt/mammoclass/). It consists of a web tool that allows users to enter a small set of BI-RADS descriptors, and produces a probability of the patient having a malignant or benign finding. MammoClass allows radiologists to click and select specific BI-RADS descriptors values, which are used to produce a prediction. However, several physicians prefer to type in or dictate full mammography reports instead

2372-9198/16 $31.00 © 2016 IEEE
DOI 10.1109/CBMS.2016.25

of entering individual descriptors in a form. Very often, the BI-RADS descriptors are available in the free-text medical reports, and are not annotated in a structured format. Moreover, some descriptors that can be relevant for posterior analysis are not annotated. In this work, we extend MammoClass with two features: (1) an interface where physicians can enter a free-text medical report, and (2) an interface for dictating the medical report, which will result in a free-text. Both free-text medical reports are parsed, the extracted BI-RADS descriptors fill up the original MammoClass form, and populate a database. The MammoClass classifier then makes a probabilistic prediction using those descriptor values. Before prediction, the form can be checked and corrected by the user. If the parser cannot extract some descriptor, the system emits an alert and suggests that the user manually chooses a value for that variable. The speech-to-text version of MammoClass can be publicly accessed at http://mammoclass.dcc.fc.up.pt/.

Medical reports of mammography findings collected at Centro Hospitalar São João, Porto, Portugal (written in Portuguese), were used as a testbed. Results show that both typed and dictated reports extract the same BI-RADS descriptors, despite the not so great performance of the speech-to-text adopted, to the Portuguese language: Google's Web Speech API (Application Programming Interface, https://dvcs.w3.org/hg/speech-api/raw-file/tip/speechapi.html).

This paper is organized as follows. Section II discusses about some speech recognition software. Section III gives an overview of BI-RADS terminology. Section IV describes the architecture of our system, and the MammoClass tool. Section V shows and discusses the performance of the chosen speech-to-text software using various devices and different voices for several BI-RADS attributes. In this Section, we also assess our interface by entering real Portuguese medical reports. Finally, we conclude and present perspectives of future work.

## II. Speech Recognition Software

Speech-to-text software is a type of software that effectively takes audio content and transcribes it into written words in a word processor or other display destination. This type of speech recognition software is extremely valuable to anyone who needs to generate written content without too much manual typing. It is also useful for people with disabilities that prevent them from using a keyboard.

For this work, we searched for various speech-to-text tools in order to choose the most suitable to our purposes: being able to recognize Portuguese words, particularly terms belonging to the BI-RADS lexicon, with a reasonable accuracy, and preferably, free. We started our research by looking at Free Voice to Text (http://download.cnet.com/Free-Voice-to-Text/3000-7239_4-76115951.html). This software can be used to send emails and documents just dictating. As the name implies it is a free tool and supports the following languages: English, Spanish, French and Japanese. Next was Talking Desktop (http://voice-recognition-software-review.toptenreviews.com/talkingdesktop-review.html). In addition to making text recognition, this software has functions to recognize dictated text about weather conditions to emit meteorological warnings. This tool is proprietary and is priced at $47. Its supported languages are English, Spanish, French and German. The next software is Dragon Naturally Speaking Home (and Premium version) (http://www.nuance.com/for-business/by-product/dragon/product-resources/edition-comparison/index.htm). After performing some basic testing, it proved to be quite accurate. It has a very functional and user-friendly interface. However, it only supports the English language. Its home edition is priced from $99, premium edition is priced at $199 and the professional edition at $599. Freesr Speech Recognition (http://freesr.org) has the ability to recognize multiple dictated texts. It assigns a number to each of the dictation interface window and allows dictated text for each one of them. It is proprietary and prices are only provided under request. It has a trial version, but it only supports the English language. Simon (https://simon.kde.org) is a free open source software available for windows and linux but it only supports the English language. All of these examples require that the user installs some software. Others allow the access via web browser or via Application Programming Interfaces (API). One of them is the Web Speech API (https://dvcs.w3.org/hg/speech-api/raw-file/tip/speechapi.html). Its Google API, which allows the programmer to obtain a translation of voice to text, has the advantage of attending our criteria: it supports the Portuguese language, and it is free. Finally, we tested Voice Note (https://voicenote.in) that is a free extension for Google Chrome. This also supports the Portuguese language and it is free. Table I summarizes the various software according to two of our selected criteria: supported languages and price (presented in US dollars), and supported platforms.

According to Table I, the only tools that meet our requirements (free and capable of understanding Portuguese words) are the Web Speech API and Voice Note. We then performed various tests with Portuguese sentences using both tools and concluded that the results for both were very similar. This fact and the fact that Voice Note is an extension for Google Chrome led us to believe that VoiceNote was built using the Web Speech API (unfortunately, the documentation is not clear about this issue). We then gave preference to the Web Speech API.

## III. Breast Cancer Terminology and Parsing of Medical Reports

Breast screening is the regular examination of a woman's breast to find breast cancer earlier. The sole exam approved for this purpose is mammography. The Breast Imaging Reporting and Data System (BI-RADS), developed by the American College of Radiology (ACR), standardized the terminology used in mammography reports [11]. The BI-RADS lexicon (some terms are shown in Table II) was created to improve the consistency of descriptive terms used in the analysis and assessment of lesion features [14]. It is composed by groups of descriptors, namely, mass shape, mass margins, mass density,

TABLE I
COMPARISON OF DIFFERENT TOOLS

| Software | Free | Price (US$) | Languages | Platform |
|---|---|---|---|---|
| Free Voice to Text | Yes | 0 | English, Spanish, French and Japanese | Windows |
| Talking Desktop | No | 47 | English, Spanish, French and German | Windows |
| Dragon Naturally Speaking Home | No | 99 | English | Windows |
| Freesr Speech Recognition | Trial | NA | English | Windows |
| Simon | Yes | 0 | English | Linux, Windows |
| Web Speech API | Yes | 0 | Portuguese and many more | All |
| Voice Note | Yes | 0 | Portuguese and many more | All |

calcification distribution, calcification morphology, associated findings and special cases. Despite the existence of the BI-RADS lexicon, there are still medical reports with a variety of concepts that are not uniformly described. Some physicians use different words to refer to a single concept. Some lexicons may be described in an ambiguous way or refer to more than one concept. In this work, we rely on the help of a radiologist to understand how the medical reports were written as well as the synonyms used to represent some BI-RADS features. All the BI-RADS terms were translated into Portuguese [15].

The original MammoClass interface allowed inputing of a specific set of BI-RADS descriptors values. As physicians give preference to enter free text or dictate their reports, we extended the interface to allow the two choices for entering reports. Because the MammoClass classifier needs specific BI-RADS descriptors, we added a parser to the interface in order to extract those descriptors from the texts. We used a parser that extracts BI-RADS descriptors from Portuguese texts by Cunha et al. [15]. The parser, written in Perl, uses regular grammar expressions and outputs results to a CSV (comma-separated value) format with boolean values, where 1 means that the descriptor was found in the text, and 0 means it was not found. Experiments with this parser has shown concordance with a specialist in over 84.5% of the extracted features. The parser also correctly classified 52.1% of the cases where there was disagreement between the parser and the specialist [15].

In order to convert words and sentences from the medical reports into concepts we cannot rely only on the BI-RADS lexicon. An example is the concept $skinlesion$, which is captured by the presence of the words $lesion$ and $skin$ separated, but with a relative proximity between the two words. Due to the fact that physicians often use different words to describe the same term, a group of words was created [15] in order to capture the intended term. As an example, to classify the associated finding $AxillaryAdenopathy$, the words $Adenomegaly$, $PositiveAxilla$ and $AxillaryNode$ were defined.

## IV. User Interface

As aforementioned, the original interface used by MammoClass only allowed the user to enter specific BI-RADS descriptors through a form. A new interface was built to allow the user to enter free text in a text box and to dictate the mammography report. The construction was based on Bootstrap (http://getbootstrap.com/), a framework to build web pages, that provides a modern look and can be comfortably used in mobile devices in addition to computers. The speech recognition module recognizes the dictated text, which is parsed. If the parser extracts any BI-RADS desriptor from the recognized text, the interface automatically fills up the original form. If any relevant variable value is not recognized or the user does not report it, the interface emits an alert. The text box allows the user to manually enter free texts. The free text is also parsed and the interface behaves in the same way as when text is recognized: it fills up the form or emits an alert in case some relevant variable value is missing.

Figure 1 shows a screenshot of the main interface for dictating texts. The MammoClass form is shown in Figure 2.

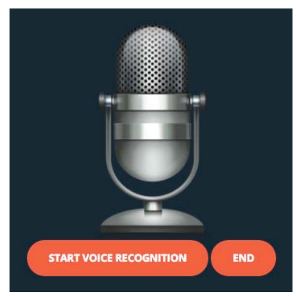


Fig. 1. Interface of Speech to Text

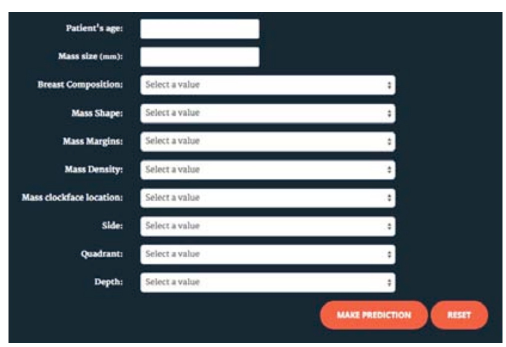


Fig. 2. Form of MammoClass

The system architecture is a client-server web-based and written in Javascript (client side) and PHP (server side). When the user (client) clicks on the button “Start Voice Recognition”

(Figure 1), the Web Speech API is called and the voice is captured. The user needs to click the button to authorize recording. When the user presses the button "End", the Web Speech API closes capturing and returns the recognized sentence. This sentence is then sent to the server. When the server receives the phrase, it calls the parser that extracts the relevant BI-RADS information from the text. The server then sends a table with this information to the client. In the client side, the JavaScript fills in the fields (Figure 2) with the extracted information. If a value for a field is not extracted by the parser, a warning message is placed besides that field in the form.

## V. EVALUATION

BI-RADS terms were translated to Portuguese and some examples are shown in Table II. In total, we have 86 different terms that appear in the BI-RADS lexicon. With these 86 terms we tested the interface by dictating all of them (Experiment 1). We also performed experiments with 67 complete medical reports of mammography results collected between 2008 and 2009, from Centro Hospitalar São João (CHSJ), Porto, Portugal. This work is approved by the Committee of Ethics of the CHSJ (reference CES 42-15). This is Experiment 2.

In order to perform experiments we created a simple interface only to dictate the terms and promptly get the result returned by the Web Speech API.

*a) Experiment 1: testing individual BI-RADS terms:* For experiment 1, we classify each returned result as: *correct* (C) if the original term and the recognized term are exactly alike; *Almost correct* (AC) if the original term and the term returned by the Web Speech API are almost identical; and *Incorrect* (I) if the original term and the term returned by the API are completely different. An almost correct term refers only to words that differ in gender, number or degree from the original term or to different spellings. An example of this disagreement can be seen, for example, in the term *distorcao arquitectural* wherein the Web Speech API returned *distorcao arquitetural*. As we see the only difference is the lack of character *c* which was removed in the new orthographic agreement among Portuguese speaking countries. *Almost correct* terms are acceptable and can be automatically corrected. *Incorrect* terms cannot be corrected and are sources of errors.

This first set of experiments were performed by four people, two male and two female. Each of these people performed the dictating test with 3 different devices: (1) a laptop with an external microphone NGS brand, (2) the same laptop with built-in microphone, and (3) a Smartphone.

We identify the male persons as A and B and the female persons as C and D.

We then evaluated the performance of the Speech API according to two dimensions: performance per person and performance per type of equipment.

Table III shows the performance of the Web Speech API for each individual experiment and the average per type of equipment used, in percentages.

As an example, person A using the laptop with an external microphone produced a correct rate of 67.4%, an almost correct rate of 8.2% and an error rate of 24.4%. Joining the rates for correct and almost correct (assuming we automatically correct the almost correct terms) we reach a total correct rate of 75.6%. For this person with the laptop with built-in microphone the Web Speech API correctly classified 68.6% of the terms, classified to almost correct 9.3% of the terms and classified as incorrect 22.1% of the terms. If we join the classes of correct and almost correct we can obtain a value of 77.9% of the terms being correct. Finally the test of person A done with the Smartphone results in 69.8% of the terms were correctly classified, 8.1% of the terms were classified to almost correct and 22.1% of the terms as classified to incorrect. If we join the classes of correct and almost correct we can obtain a value of 77.9% of the terms being correct.

TABLE III
PERFORMANCE PER EXPERIMENT

| **Device** | Person | **C(%)** | **AC(%)** | **I(%)** | **C+AC(%)** |
|---|---|---|---|---|---|
| Laptop with ext micro | A | 67.4 | 8.2 | 24.4 | 75.6 |
| | B | 77.9 | 5.8 | 16.3 | 83.7 |
| | C | 68.6 | 9.3 | 22.1 | 77.9 |
| | D | 66.3 | 7.0 | 26.7 | 73.3 |
| | Avg. | 70.1 | 7.5 | 22.4 | 77.6 |
| Laptop with int micro | A | 68.6 | 9.3 | 22.1 | 77.9 |
| | B | 74.4 | 4.7 | 20.9 | 79.1 |
| | C | 67.4 | 4.7 | 27.9 | 72.1 |
| | D | 66.3 | 3.5 | 30.2 | 69.8 |
| | Avg. | 69.2 | 5.5 | 25.3 | 74.7 |
| Smartphone | A | 69.8 | 8.1 | 22.1 | 77.9 |
| | B | 74.4 | 7.0 | 18.6 | 81.4 |
| | C | 70.9 | 7.0 | 22.1 | 77.9 |
| | D | 60.5 | 5.8 | 33.7 | 66.3 |
| | Avg. | 68.9 | 7.0 | 24.1 | 75.9 |

TABLE IV
AVERAGES PER PERSON

| **Person** | **C(%)** | **AC(%)** | **I(%)** | **C+AC(%)** |
|---|---|---|---|---|
| A | 68.6 | 8.5 | 22.9 | 77.1 |
| B | 75.6 | 5.8 | 18.6 | 81.4 |
| C | 69.0 | 7.0 | 24.0 | 76.0 |
| D | 64.3 | 5.5 | 30.2 | 69.8 |

For each type of equipment used, we calculated the average performance for the four people (line labeled "Avg." in Table III). We also calculated the average performance per person, as shown in Table IV. For person A we have an average of 68.6% hits, an average of 8.5% almost correct classifications and an average of 22.9% incorrect classifications. If we join the averages of correct and almost correct we can obtain a value of 77.1% of the terms being correct.

Values in this range are also reported in other works where radiology reports are dictated. We did not compare the Google API results with other software, because no other had the support to understand Portuguese or was freely available.

*b) Experiment 2: testing medical reports:* In experiment 2, we compared the output of our parser when dictating the text and when copying and pasting the medical report on the interface's textbox. The dictation was performed by just one person. Examples of reports (in Portuguese) can be found in Figure 3. The first report talks about a mass of size 2

TABLE II
BI-RADS TERMS AND THEIR TRANSLATION TO PORTUGUESE

| | | |
|---|---|---|
| mass shape | round<br>oval<br>lobular<br>irregular | arredondada, redonda<br>oval, ovóide, alongada, ovalar<br>lobular, Polilobular<br>irregular |
| mass margins | circumscribed<br>microlobular<br>obscure<br>indistinct<br>spiculated | circunscrita, bem definida, bem delimitada, regular<br>microlobular<br>obscura, obscurecida<br>indistinta, imprecisa, indefinida, mal definida<br>espiculada |
| mass density | high<br>equal<br>low | alta, elevada densidade<br>igual, isodensa, homogênea<br>baixa, tênue |
| breast density | predominantly fatty<br>scattered fibroglandular<br>heterogeneously dense<br>extremely dense | contém gordura<br>fibroglandular<br>heterogênea<br>densa, muito densa, densidade alta da mama |
| calcification morphology | skin calcifications<br>vascular calcifications<br>coarse calcifications<br>large rod-like calcifications<br>round calcifications<br>lucent-centered calcifications<br>eggshell calcifications<br>milk of calcium calcifications<br>suture calcifications<br>dystrophic<br>punctate<br>amorphous<br>pleomorphic<br>fine linear branching | calcificações dérmicas, calcificações pele<br>calcificações vasculares<br>calcificações grosseiras, pipoca<br>calcificações em forma de bastonete<br>calcificações redondas<br>calcificações com centro lucente<br>calcificações em casca de ovo<br>calcificações ductais ou leite cálcio<br>calcificações cicatriciais<br>calcificações distróficas<br>calcificações punctiformes<br>calcificações amórficas ou indistintas<br>calcificações pleomórficas ou heterogéneas<br>calcificações finas ou lineares |
| calcification distribution | clustered<br>linear<br>segmental<br>regional<br>scattered | agrupadas, com vários núcleos<br>microcalcificações lineares<br>segmentar<br>regional, área extensa<br>difusa, dispersas morfologicamente |
| special cases | architectural distortion<br>solitary dilated duct<br>intrammary lymph node<br>assymetric breast tissue | distorção arquitectural, desorganização arquitectural<br>dilatação ductal, ducto dilatado, carcinoma ductal<br>gânglio linfático intramamário, gânglio mamário<br>densidade assimétrica, assimetria mamária |
| associated findings | skin retraction<br>trabecular thickening<br>nipple retraction<br>skin lesion<br>axillary adenopathy | retracção cutânea<br>espessamento trabecular<br>retracção do mamilo<br>lesão na pele, cutânea, ulceração cutânea<br>adenopatia axilar, adenomegalia axilar, axila positiva, gânglio axilar suspeito |

centimeters in the left superior quadrant of the breast with findings that are suspicious of malignancy. The final BI-RADS category assigned is 5. The second report talks about a breast with predominantly fatty density, with normal distribution of fibroglandular elements, trauma well-known to the right, stroma distortion located at the right breast with 30 millimeters of size.

Each text was pasted to the interface's text box, and dictated. For all 67 reports, the parser managed to extract the same relevant BI-RADS terms, with exception of the mass sizes. While the pasted texts have abbreviations of the units (cm for centimeters, "centímetros", in Portuguese, and mm for millimeters, "milímetros", in Portuguese), the dictated text was spelled out. As the parser is prepared to understand the abbreviated units, these sizes were only extracted from the pasted texts. However, it is trivial to change our parser to incorporate the whole spelling of size units or other kinds of units into the parser. Figure 4 shows the result of the Web Speech API after dictating the two texts of Figure 3.

1. *Nódulo com cerca de 2cm, QSE da mama esquerda, com achados suspeitos de malignidade. Achados imagiológicos muito sugestivos de malignidade - Bi-Rads - 5*

2. *Glândulas mamárias predominantemente adiposas, com normal distribuição dos elementos fibro-glandulares existentes. Antecedentes de Trauma conhecido á direita Distorção do estoma. localizada na mama Direita quadrante Superior-Externo com 30mm de tamanho. Alterações com suspeição de malignidade intermédia - Bi-Rads - 4b. Aconselhado efectuação de Microbiópsia ecoguiada.*

Fig. 3. Medical Report Examples

As expected for this kind of tool and considering that it is performing the recognition for the Portuguese language (not yet very explored in speech-to-text recognition), the recognized text has some sections that do not make sense. However, the main BI-RADS terms, relevant to fill up the

1. *nódulo com cerca de 2 centímetros psn esquerda concha suspeitas de malignidade achados imagiologicos muitos estilos maligna hi5*

2. *planos marisco de nascimento de casas com armas distribuição dos elementos fibroglandulares acidente traumático cidade dos santos toma da madeira quadrante superior externo com 30 milímetros tem alterações conceição idade média ps4 de concentração microbiopsia ecoguiada*

Fig. 4. Recognized Texts

form for posterior classification of the finding, are captured from the dictated text.

## VI. Discussion

From experiment 1, we can directly conclude that the Web Speech API can be sensitive to the type of voice (entonation or way of speaking). For example, person B caused the API to result in the lowest error rate (average 18.6% as shown in Table IV. Although person A has a higher rate of almost correct terms (8.5%, which can be corrected automatically), his error rate is the highest (22.9%). If we use the last column of Table IV as a performance metric for the Web Speech API, person B would be the best one for dictating tests. Note that this Speech API does not learn new voice patterns. On the other hand, if we measure the performance of the Speech API according to the type of equipment we use, clearly the use of an external microphone gives better results, followed by the smartphone, and by the use of the laptop with an internal microphone.

From experiment 2, we learned 2 lessons:

- Speech interfaces for long sentences in Portuguese need to be improved. The one we tested is very sensitive to the speech speed and is not well trained to the language itself.
- Although the recognized text sometimes differs from the original written report, the most relevant BI-RADS terms are still recognized.

## VII. Conclusions and Future Work

Applications interfaces are very important to allow full adoption of new technologies in the health domain. In this work, we explored the domain of breast cancer and provided to the user an interface where medical reports can be dictated as opposite to input in forms or textboxes. As our tool is used to perform classification of breast cancer findings, this feature will promote a wider use of the interface allowing to populate a database of new cases that can be used later to refine our classifier, and providing an easier way of testing new cases supporting medical decision. However, the API used in our study is somewhat limited. Besides, Google API would not be a suitable choice for entering sensitive patient data available in actual hospital environments. We would like to design and implement our own tools for recognizing Portuguese terms, which could be independent of voice type or entonation, and that could be trained only on the subset of words used in the area of breast cancer.


## Acknowledgments

This work is supported by the ERDF – European Regional Development Fund through the Operational Programme for Competitiveness and Internationalisation – COMPETE 2020 Programme, by National Funds through the FCT – Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) within project POCI-01-0145-FEDER-006961, and Project "NORTE-01-0145-FEDER-000016", financed by the North Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, and through the European Regional Development Fund (ERDF).


## References


[1] H. P. Kang, S. J. Sirintrapun, R. J. Nestler, and A. V. Parwani, "Experience with voice recognition in surgical pathology at a large academic multi-institutional center," *American Journal of Clinical Pathology*, vol. 133, no. 1, pp. 156–159, 2010.

[2] W. H. Henricks, K. Roumina, and B. E. Skilton, "The utility and cost effectiveness of voice recognition technology in surgical pathology," *Mod Pathol*, vol. 15, 2002.

[3] P.-Y. Yen and S. Bakken, "Review of health information technology usability study methodologies," *Journal of the American Medical Informatics Association*, vol. 19, no. 3, pp. 413–422, 2012.

[4] J. du Toit, R. Hattingh, and R. Pitcher, "The accuracy of radiology speech recognition reports in a multilingual south african teaching hospital," *BMC Medical Imaging*, vol. 15, no. 1, pp. 1–5, 2015. [Online]. Available: http://dx.doi.org/10.1186/s12880-015-0048-1

[5] S. Basma, B. Lord, L. M. Jacks, M. Risk, and S. A. M., "Error rates in breast imaging reports: comparison of automatic speech recognition and dictation transcription." *AJR Am J Roentgenol*, vol. 197, pp. 923–927, 2011.

[6] R. Hoyt and A. Yoshihashi, "Lessons learned from implementation of voice recognition for documentation in the military electronic health record system," *Perspectives in Health Information Management*, no. 7(Winter):1e, 2010.

[7] S. McGurk, K. Brauer, T. V. Macfarlane, and K. A. Duncan, "The effect of voice recognition software on comparative error rates in radiology reports," *The British Journal of Radiology*, vol. 81, pp. 767–770, 2008.

[8] I. Hammana, L. Lepanto, T. Poder, and M. S. Bellemare, C. Ly, "Speech recognition in the radiology department: a systematic review," *HIM J.*, vol. 44, no. 2, pp. 4–10, 2015.

[9] R. Patel, B. Greenberg, S. Montner, A. Funaki, C. Straus, S. Zangan, and H. MacMahon, "Reduction of voice recognition errors in radiological dictation: Effects of systematic individual feedback," http://clinicaleffectiveness.uchicago.edu/files/2013/07/Reduction-of-Voice-Recognition-Errors-in-Radiological-Dictation-Effects-of-Systematic-Individual-Feedback.pdf, accessed in Feb 2016.

[10] D. Sonntag, M. Weber, A. Cavallaro, and M. Hammon, "Integrating digital pens in breast imaging for instant knowledge acquisition," *AI Magazine*, vol. 35, no. 1, pp. 26–37, 2014. [Online]. Available: http://www.aaai.org/ojs/index.php/aimagazine/article/view/2501

[11] E. A. Sickles, C. J. D'Orsi, L. W. Bassett, and et al., "ACR BI-RADS® Mammography," in *In: ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System*, 2013.

[12] P. M. Ferreira, N. A. Fonseca, I. Dutra, R. Woods, and E. Burnside, "Predicting malignancy from mammography findings and image-guided core biopsies," *International Journal of Data Mining and Bioinformatics*, vol. 11, p. 257–276, March 2015.

[13] J. Platt, "Machines using sequential minimal optimization," in *Advances in Kernel Methods - Support Vector Learning*, B. Schoelkopf, C. Burges, and A. Smola, Eds. MIT Press, 1998.

[14] L. Liberman, A. F. Abramson, F. B. Squires, J. R. Glassman, E. A. Morris, and D. D. Dershaw, "The breast imaging reporting and data system: positive predictive value of mammographic features and final assessment categories," *AJR Am J Roentgenol.*, vol. 171, no. 1, pp. 35–40, 1998.

[15] H. Nassif, F. Cunha, I. C. Moreira, R. Cruz-Correia, E. Sousa, D. Page, E. S. Burnside, and I. de Castro Dutra, "Extracting bi-rads features from portuguese clinical texts," in *IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2012*, 2012, pp. 1–4.