

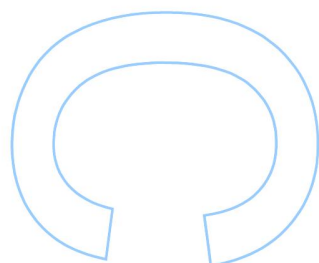
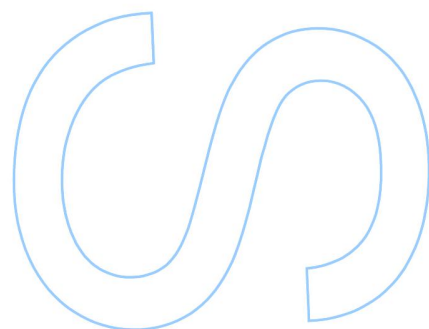
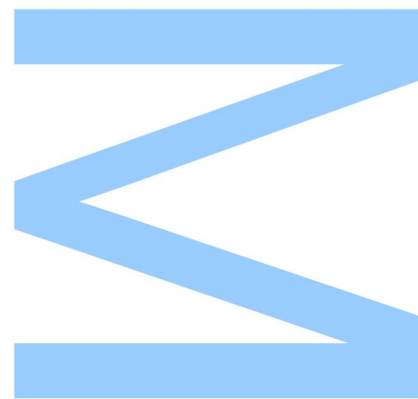
Implementação de uma ferramenta de suporte à entrada de texto em Português, escrito e falado, para o sistema de apoio à decisão clínica MammoClass

Ricardo Daniel de Sousa Rocha

Mestrado em Ciência de Computadores
Departamento de Ciência de Computadores
2016

Orientador

Inês de Castro Dutra, Professora Auxiliar
Faculdade de Ciências da Universidade do Porto

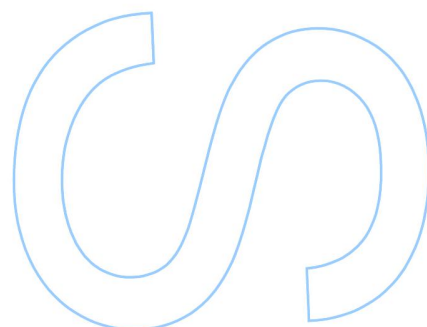
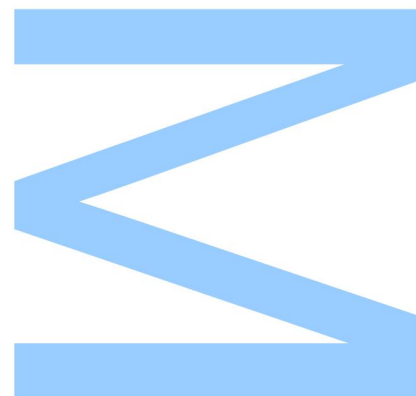




Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, ____ / ____ / ____



Para os meus pais. Para a minha namorada. Em memória do meu avô.

Agradecimentos

Quero agradecer à minha orientadora, Professora Doutora Inês Dutra, por todos os ensinamentos que me passou, toda a paciência que teve para as minhas dúvidas, por todo o apoio e amizade que me deu ao longo deste projeto.

Quero agradecer ao meu pai, António, e à minha mãe, Celeste por todo o suporte que me deram. Foram eles que me trouxeram ao mundo, mas mais que isso permitiram que eu chegasse aqui. Sempre me proporcionaram tudo o que precisei e ensinaram-me a lutar pelo que quero. São os pais que toda a gente sonha ter.

Quero agradecer à minha namorada Jordana, toda a força que me dá, todo o apoio, amizade, amor, e confiança que me transmite. Quando eu pensava que não era capaz, sempre fez com que eu não desistisse. Ainda me lembro da conversa que tivemos no parque da cidade há 5 anos atrás, talvez sem ela nunca tivesse chegado ao ponto de escrever uma dissertação.

Sem individualizar para não estender muito este texto, quero agradecer a todos os familiares e amigos que fazem parte da minha vida de forma ativa. Cada um teve a sua contribuição no meu percurso, de forma maior ou menor, mas igualmente importante.

Quero agradecer ao professor Fernando Silva, por todo o apoio e disponibilidade demonstrada ao longo da formação académica, e um obrigado a todos os outros docentes por todos os ensinamentos que me transmitiram ao longo dos anos que passei nesta instituição. Um obrigado ao staff do Departamento de Ciência de Computadores por permitirem que tenhamos as condições necessárias para aprender e por estarem sempre prontos a ajudar.

Quero também agradecer o suporte financeiro realizado neste projeto pela *ERDF - European Regional Development Fund* através do programa operacional competitividade e operacionalização -*COMPETE 2020 Programme* e por fundos nacionais através da Fundação para a Ciência e Tecnologia (FCT) com o projeto *POCI-01-0145-FEDER-006961*.

Por último mas não menos importante, quero agradecer ao meu avô, Joaquim dos Santos Sousa. Infelizmente já não está entre nós fisicamente, contudo sei que esteve e está sempre comigo.

Obrigado...

*Ricardo Daniel de Sousa Rocha
Porto, Setembro de 2016*

Resumo

Ferramentas de *Speech-to-Text* (STT) têm sido utilizadas com sucesso em alguns domínios, especialmente em ambientes fechados, onde o ruído é reduzido, levando a que o texto ditado possa ser convertido em texto escrito com uma qualidade razoável. Na área médica, STT é muito usado, especialmente para ditar relatórios de imagem. A qualidade do texto reconhecido a partir do discurso nem sempre é de muito boa qualidade. No entanto, em algumas áreas, não é relevante reconhecer corretamente o texto completo que foi ditado. Por exemplo, na área de radiologia da mama, as imagens são anotadas de acordo com um léxico padrão, o *Breast Imaging Reporting and Data System* (BI-RADS). Este léxico define uma coleção de termos padrão que podem ser anotados numa mamografia (Raio-X da mama). Se estamos interessados apenas nos dados estruturados relacionados com a mamografia, e o STT é capaz de reconhecer esses termos, não há problema se algumas palavras no texto forem reconhecidas incorretamente. Dados estruturados, ao contrário de textos em linguagem natural, são úteis para a análise de dados e construção de modelos de previsão.

Neste trabalho, combinamos várias tecnologias para implementar uma ferramenta para processamento do discurso no domínio da mamografia, e extração de termos BI-RADS deste discurso. Esta é especializada para o Português. Esta ferramenta é integrada ao MammoClass, um aplicativo baseado na web que pode prever a probabilidade de um achado numa mamografia ser benigno ou maligno. Além de ser capaz de processar texto ditado, também pode processar relatórios médicos escritos. A nova aplicação MammoClass, resultado do nosso trabalho, permite três tipos de entrada de dados: texto ditado, texto escrito e baseada em formulário. As duas primeiras opções são o resultado deste trabalho. Uma vez que o utilizador insere texto usando uma dessas duas opções, um analisador do Português extrai os termos BI-RADS do texto e

preenche automaticamente o formulário. O utilizador pode então ter a oportunidade de modificar qualquer campo do formulário.

O utilizador pode escolher entre duas ferramentas STT diferentes: *Web Speech API* ou Julius / coruja. Mostramos que ambos têm desempenho semelhante em relação à qualidade dos termos extraídos, e são bem adequados para a tarefa. A *Web Speech API* tem uma desvantagem, uma vez que não garante a confidencialidade dos dados.

No que concerne à performance de termos BI-RADS as duas ferramentas obtêm uma média de acerto de aproximadamente 80%.

Palavras-chave: Conversão de voz em texto, interfaces web, sistemas de apoio à decisão clínica

Abstract

Speech-to-Text (STT) tools have been used with success in some domains, specially in closed environments, where the noise is reduced, and dictated text can be converted to written text with a reasonable quality. In the medical area, STT is very much used, specially to dictate image reports. The quality of the recognized speech is not always of very good quality. However, in some areas, it is not relevant to correctly recognize the full dictated text. For example, in the area of breast radiology, images are annotated according to a standard lexicon, the Breast Imaging Reporting and Data System (BI-RADS). This lexicon defines a collection of standard terms that can be annotated from a mammography (X-Ray of the breast). If we are only interested in the structured data related to the mammography, and the STT is capable of recognizing those terms, it will be OK if some words in the text are incorrectly recognized. Structured data, as opposed to natural language texts, are useful to data analysis and to build prediction models.

In this work, we combine several technologies to implement a tool that can process speech in the mammography domain, and can extract BI-RADS terms from the speech. This is specialized to the Portuguese language. This tool is integrated to MammoClass, a web-based application that can predict the probability of a finding in a mammography be benign or malignant. Besides being able to process dictated text, it can also process written medical reports. The new MammoClass application, result of our work, can allow three types of data input: dictated text, written text and form-based. The first two features are the result of this work. Once the user enters text using one of these two options, a parser of the Portuguese language extracts the BI-RADS terms from the input, and automatically fills up the form. The user can then have a chance to modify any form field.

The user can choose between two different STT tools: Web Speech API or Julius/coruja. We show that both have similar performance regarding quality of the extracted terms, and are well suited to the task. The Web Speech API has a disadvantage as it does not guarantee data confidentiality.

The performance of both tools regarding percentage of correctly extracted features is around 80%.

Keywords: Speech recognition, Speech-to-Text, web interfaces, clinical decision support

Conteúdo

Resumo	v
Abstract	vii
Índice de Tabelas	xiii
Índice de Figuras	xvi
Glossário	xvii
1 Introdução	1
2 Fundamentos e Terminologia	5
2.1 O Cancro	5
2.2 O cancro da Mama	6
2.2.1 Mamografia	7
2.2.2 Sistema BI-RADS®	9
2.3 Arquitetura Cliente-Servidor	19
2.4 <i>Speech-To-Text</i>	20
3 Estado da Arte e Trabalhos Relacionados	23

4	Ferramentas de <i>Speech-to-Text</i>	29
4.1	<i>Web Speech API</i> ou <i>VoiceNote?</i>	31
4.2	Problemas inerentes ao uso da <i>Web Speech API</i>	32
4.3	Julius e Coruja	33
5	Speech-to-Text Interface	37
5.1	MammoClass	38
5.2	<i>Parser</i> extrator de descritores BI-RADS®	38
5.3	Interface Criada	43
5.3.1	Interface Inicial	43
5.3.2	Interface para <i>Speech-To-Text</i>	46
5.3.3	Interface para digitar texto	48
5.3.4	Interface com o formulários de descritores BI-RADS®	50
5.3.5	Interface de informação relativa ao projeto	54
5.4	Explicação e Esquematização da nova Ferramenta	55
6	Metodologia Experimental	61
6.1	Materiais e Métodos	61
6.2	Experiências	63
6.2.1	Experiência 1 - Teste aos termos BI-RADS®	63
6.2.2	Experiência 2 - Teste aos relatórios médicos	64
7	Resultados e Discussão	65
7.1	Experiência 1 - Teste aos termos BI-RADS®	65
7.2	Experiência 2 - Teste aos relatórios médicos	68
7.3	Discussão	69

8	Conclusão	71
	Bibliografia	73
9	Anexos	77
9.1	Termos BI-RADS [®] extraídos pela <i>Web Speech API</i>	77
9.2	Instalação do Julius e Coruja num servidor	88

Índice de Tabelas

2.1	Léxico BI-RADS Forma, adaptado de Cunha (2011)	10
2.2	Léxico BI-RADS Margens, adaptado de Cunha (2011)	11
2.3	Léxico BI-RADS Densidade, adaptado de Cunha (2011)	11
2.4	Léxico BI-RADS Morfologia (Tipicamente Benignas), adaptado de Cunha (2011)	12
2.5	Léxico BI-RADS Morfologia (Probabilidade de Malignidade Intermédia), adaptado de Cunha (2011)	12
2.6	Léxico BI-RADS Morfologia (Alta Probabilidade de Malignidade), adaptado de Cunha (2011)	13
2.7	Léxico BI-RADS Distribuição das Calcificações, adaptado de Cunha (2011)	13
2.8	Léxico BI-RADS Distorção Arquitetural, adaptado de Cunha (2011)	13
2.9	Léxico BI-RADS Casos Especiais, adaptado de Cunha (2011)	14
2.10	Léxico BI-RADS Achados Associados, adaptado de Cunha (2011)	15
2.11	Categorias BI-RADS [®] , adaptado de Cunha (2011)	15
4.1	Comparação dos <i>Software Speech-To-Text</i>	31
4.2	Comparação entre <i>Web Speech API</i> e <i>VoiceNote</i> - Exemplo 1	32
4.3	Comparação entre <i>Web Speech API</i> e <i>VoiceNote</i> - Exemplo 2	32
4.4	Comparação entre <i>Web Speech API</i> e <i>VoiceNote</i> - Exemplo 3	32

5.1	Termos BI-RADS [®] e a sua tradução para Português	42
7.1	Performance por pessoas	67
7.2	Médias por pessoas	67

Índice de Figuras

2.1	Diferenças entre tecidos mamários presentes numa mamografia. Retirado de Bontrager (2006)	8
2.2	Evolução da mama segundo a idade. Retirado de Dores (2014)	9
2.3	Léxico BI-RADS [®] adaptado de Nassif <i>et al.</i> (2009)	10
2.4	Imagens referentes a duas mamografias distintas. A mamografia da esquerda apresenta uma mama normal, em que as áreas mais densas (brancas) são os canais galactóforos. A mamografia da direita, por sua vez, apresenta uma área branca densa (canto inferior direito da imagem) que indica a presença de um tumor, retirado de Ferreira (2010)	16
2.5	Anatomia de uma mama saudável, retirado de Ferreira (2010)	17
2.6	Representação de uma arquitetura Cliente-Servidor	20
4.1	Fluxo dos dados para incluir o <i>Julius</i> e o <i>Coruja</i> no MammoClass	36
5.1	Interface do Mammoclass original	39
5.2	Gramática para extração do conceito "Forma"	41
5.3	Esquema da interface do MammoClass	43
5.4	Interface apresentada aquando da visita à pagina do MammoClass - Computador e <i>Tablet</i>	44
5.5	Interface apresentada aquando da visita à pagina do MammoClass - <i>Smartphone</i>	45

5.6	Menu presente sempre no topo da interface - Computador e <i>Tablet</i> . . .	45
5.7	Menu acessível sempre no topo da interface - <i>Smartphone</i>	45
5.8	Botões de navegação fácil presentes na interface inicial	46
5.9	Interface responsável pela componente <i>Speech-To-Text</i> - Computador e <i>Tablet</i>	47
5.10	Interface responsável pela componente <i>Speech-To-Text</i> - <i>Smartphone</i> . .	47
5.11	<i>Pop-up</i> do <i>Google Chrome</i> para permissão de recolha de voz	48
5.12	Interface responsável pela componente <i>Speech-To-Text</i> aquando gravação - Computador e <i>Tablet</i>	49
5.13	Interface responsável pela componente <i>Speech-To-Text</i> aquando gravação - <i>Smartphone</i>	49
5.14	Interface responsável pela componente de texto ditado - Computador e <i>Tablet</i>	50
5.15	Interface responsável pela componente de texto ditado - <i>Smartphone</i> . .	51
5.16	Interface com o formulários de descritores BI-RADS [®] - Computador e <i>Tablet</i>	51
5.17	Interface com o formulário de descritores BI-RADS [®] - <i>Smartphone</i> . .	52
5.18	Aviso emitido quando formulário está incompleto.	53
5.19	Interface do formulário após uma predição	53
5.20	Interface do formulário com avisos despoletados pelo <i>parser</i> - Computador e <i>Tablet</i>	54
5.21	Interface do formulário com avisos despoletados pelo <i>parser</i> - <i>Smartphone</i>	55
5.22	Interface relacionada com a informações sobre o projeto - Computador e <i>Smartphone</i>	56
5.23	Interface relacionada com a informações sobre o projeto - <i>Smartphone</i> .	56
5.24	Gráfico de Fluxo do funcionamento da componente <i>Speech-to-text</i> . . .	57

5.25	Esquematização de toda a ferramenta	59
6.1	Interface usada para as Experiências	62
6.2	Interface usada para as Experiências após o ditado de um termo	62
7.1	Exemplos de relatórios médicos	68
7.2	Relatórios reconhecidos pela <i>Web Speech API</i>	69

Glossário

- **Application Program Interface (API)** - Uma API é um conjunto de rotinas, protocolos, métodos e ferramentas para produzir *software*. Uma API especifica de que forma as diversas componentes do *software* devem interagir. Uma boa API torna o desenvolvimento de um programa bastante mais fácil pois proporciona ao programador vários métodos que já produzem determinada funcionalidade, proporcionado assim abstração sobre a forma como são implementados certos detalhes.
- **GET** - O método GET é usado para pedir dados a um servidor. Com este método os dados do pedido são adicionados ao URL com uma sintaxe especial, ficando o seu formato da seguinte forma: `http://exemplo.com/dados?id=774` onde o parâmetro *id* é dado pelo valor 774 diretamente no URL. Este método é apropriado para pequenos parâmetros e em casos em que não haja necessidade de esconder tais parâmetros, uma vez que os parâmetros ficam visíveis no URL.
- **HTML5** - HTML significa *Hyper Text Markup Language* e tem como finalidade descrever a estrutura de uma página Web usando marcação. O HTML5 é a versão mais recente do HTML que disponibiliza ao programador novas características necessárias para a construção de aplicações modernas numa página Web. O HTML5 também tornou *standard* muitas ferramentas que os programadores vinham a usar ao longo dos anos mas que nunca tinham sido documentadas pelo comité de standardização. O uso do HTML5 permite que seja mais fácil incluir e manipular conteúdo gráfico e multimédia nas páginas Web sem ser necessário recorrer a *plugins* proprietários.
- **JavaScript** - O JavaScript é uma linguagem de programação processada pelo navegador web. Através do JavaScript podem ser criados efeitos na página Web. Permite ainda aumentar a interatividade com o utilizador.

- **PHP** - O PHP é uma linguagem de script, muito utilizada, e especialmente adequada para o desenvolvimento web e que pode ser embutida dentro do HTML. O que distingue o PHP de algo como o Javascript que opera no lado do cliente é que o código é executado no servidor, gerando então HTML que é enviado para o navegador.
- **Plugin** - Na informática, um *plugin* consiste num programa de computador usado para adicionar funções a outros programas de maior dimensão, provendo esse programa de alguma funcionalidade especial ou muito específica.
- **POST** - O método POST é usado, tal como GET para pedir dados a um servidor. Contudo no POST os parâmetros não são visíveis para o utilizador. O POST é muitas vezes usado para realizar o envio de algum ficheiro para um servidor ou para enviar formulários completos.
- **Uniform Resource Locator (URL)** - Um URL é o endereço de determinado documento ou outro recurso informático na Internet.

Alguns termos técnicos utilizados nesta dissertação foram mantidos no original em inglês. Estes estão em fonte itálico.

Capítulo 1

Introdução

O cancro é uma proliferação anormal de células (Liga Portuguesa Contra o Cancro¹). No ano de 2012, ano em que os dados mais recentes estão disponíveis, 224.147 mulheres e 2.125 homens foram diagnosticados com cancro da mama nos Estados Unidos. A estes números por si só preocupantes, pode-se acrescentar que no ano de 2012, 41.150 mulheres e 405 homens morreram devido ao cancro da mama. Ora, de acordo com Ferreira (2010), a partir de 1995 começou a verificar-se uma ligeira tendência para a diminuição da mortalidade devido à introdução de programas de rastreio, ou seja, a realização de forma periódica de exames num grupo populacional assintomático com o principal intuito de detetar o cancro da mama num estado precoce. Para esta deteção a mamografia é o exame imagiológico mais utilizado. De salientar que, de acordo com Ferreira (2010), o estudo mais antigo sobre programas de rastreio está datado entre 1963 e 1970, sendo que, desde essa data até aos dias de hoje conclui-se que houve uma redução da mortalidade por cancro da mama na ordem dos 30% quando comparado com o grupo de controlo após cinco anos. Nenhuma outra medida diagnóstica ou terapêutica levou a uma redução tão acentuada da taxa de mortalidade.

O sistema BI-RADS[®] é um sistema de classificação que foi desenvolvido para promover uma uniformização dos relatórios mamográficos, sendo que, o seu léxico é constituído por 43 descritores.

Por sua vez, a tecnologia de *Speech-To-Text* tem sido melhorada ao longo dos anos, permitindo relatórios precisos e conseqüentemente o armazenamento de mais infor-

¹<https://www.ligacontracancro.pt/>

mação de teor qualitativo. Existem trabalhos na literatura que reportam histórias de sucesso provenientes do uso de *Speech-To-Text* com o objetivo de extrair o conteúdo de textos ditados no âmbito da área médica. Contudo, alguns trabalhos na literatura não são a favor da utilização de tecnologia de *Speech-To-Text* na área da radiologia e consequentemente reportam uma taxa de erro elevada referente ao texto reconhecido. Muitos estudos salientam algumas lacunas presentes nos sistemas de *Speech-To-Text*, principalmente questões relacionadas com a quantidade de erros presentes num relatório que seja gerado por este sistema (du Toit *et al.* , 2015; Basma *et al.* , 2011; Hoyt & Yoshihashi, 2010; McGurk *et al.* , 2008; Hammana *et al.* , 2015). Contudo, dos sistemas de *Speech-To-Text* também advêm aspetos positivos, sendo o principal o pouco tempo que demora até um relatório ficar disponível. Ora, todos os estudos focam-se no texto em si, e não em palavras. As palavras poderiam ser extraídas do texto com o intuito de construir dados estruturados que possam ser usados em estudos automáticos posteriores.

Atendendo a que o radiologista usa a terminologia BI-RADS[®] para descrever os achados numa mamografia, tais descritores são bastantes úteis, já que através deles podem ser construídos modelos preditivos de forma automática e assim ajudar o clínico a suportar decisões clínicas. O MammoClass desenvolvido por Ferreira (2010) é um sistema de apoio à decisão clínica que utiliza dados estruturados (características BI-RADS[®] anotadas de imagens de mamografias) para fazer a classificação de um achado imagiológico. Tais dados estruturados podem ser inseridos manualmente no MammoClass através de um formulário.

Para esta ferramenta, o texto em si não é relevante. É suficiente que a ferramenta consiga extrair os termos BI-RADS[®] importantes contidos no relatório digitado.

O objetivo deste trabalho, é a implementação de uma ferramenta de suporte à entrada de texto em Português, escrito e falado para o MammoClass.

Apesar dos relatórios ditados para sistemas de *Speech-To-Text* terem uma elevada taxa de erro, quando é considerado o texto como um todo, estes sistemas são bastantes úteis a acelerar o processo de criar dados estruturados, já que a utilização do texto resultante terá o seu foco em palavras relevantes e não no texto como um todo.

Como resultado deste trabalho, temos as seguintes contribuições:

- uma interface web funcional para entrada de dados de mamografia e seu respectivo método computacional subjacente;
- uma apresentação no 2nd Breast Cancer Worskhop, realizado na Universidade do Porto;
- uma apresentação no IJUP 2015;
- um artigo científico publicado na Computer-Based Medical Systems (CBMS) 2016.

Esta dissertação está organizada da seguinte forma, o Capítulo 2 apresenta os fundamentos e terminologias sobre temas importantes para o melhor entendimento desta dissertação. O Capítulo 3 dá a conhecer ao leitor uma revisão literária de sistemas de *Speech-To-Text* aplicados a dados clínicos. Por sua vez o Capítulo 4 explica todas as ferramentas de *Speech-To-Text* estudadas no decorrer desta dissertação, bem como as escolhas tomadas em relação a qual ferramenta usar e também a forma de as incluir no MammoClass. O Capítulo 5 apresenta ao leitor a nova interface do MammoClass elaborada nesta dissertação bem como a arquitetura do sistema montado e a explicação da sua implementação. O Capítulo 6 descreve as experiências realizadas, métricas de avaliação e a metodologia experimental utilizada para a avaliação da nova plataforma MammoClass. No Capítulo 7 é feita a análise aos resultados obtidos no que concerne à componente *Speech-ToText* da nova versão do MammoClass desenvolvida nesta dissertação. Por último, o Capítulo 8 apresenta ao leitor as ilações retiradas da elaboração desta dissertação bem como a sugestão de futuros trabalhos que podem ser realizados para melhorar a nossa implementação.

Capítulo 2

Fundamentos e Terminologia

Este capítulo visa apresentar alguns fundamentos que achamos necessários para um melhor entendimento desta dissertação. Junto com estes fundamentos é apresentada a terminologia inerente ao tema e que achamos relevante.

2.1 O Cancro

O cancro é uma proliferação anormal de células (Liga Portuguesa Contra o Cancro¹). As células crescem e dividem-se com o intuito de formar novas células. Ao longo do seu ciclo de vida, as células envelhecem, morrem e depois são substituídas por células novas. De acordo com (Liga portuguesa Contra o Cancro) o cancro tem inicio nas células pois algumas vezes o ciclo de vida das células não corre como esperado, sendo que podem ser formadas novas células sem que o organismo necessite, podendo acontecer em simultâneo que células velhas não morram. Posto isto, esse conjunto de células extra é considerado um tumor. Os tumores podem ser benignos ou malignos. No caso de um tumor benigno este não é considerado cancro, estes são tumores que não colocam a vida em risco e regra geral podem ser removidos. De acrescentar que as células dos tumores benignos não se espalham para os tecidos em volta da células ou para outros tecidos envolventes. Por sua vez os tumores malignos são cancro. Estes tumores são mais graves que os tumores benignos e podem colocar a vida em risco. Estes tumores podem ser removidos, contudo é de ter em consideração que estes podem voltar a crescer. Ao

¹<https://www.ligacontracancro.pt/>

contrário das células dos tumores benignos, as células dos tumores malignos podem invadir e danificar os tecidos e órgãos circundantes. Além disso podem libertar-se do tumor primitivo e espalhar-se para a corrente sanguínea e provocar tumores noutros órgãos. Em termos de nomenclatura segundo (Liga portuguesa contra o cancro) o nome dado à maioria dos cancros provém do tumor inicial. Por exemplo, o cancro do pulmão tem início no pulmão e o cancro da mama tem início na mama. Segundo Siegel *et al.* (2016) o cancro é um grande problema de saúde pública em todo o mundo. Por exemplo, nos Estados Unidos da América o cancro é a segunda maior causa de morte e é expectável que supere nos próximos anos as doenças cardíacas como a doença que mais mortes causa. A título de exemplo, é possível constatar em Siegel *et al.* (2016) que o número de novos casos esperados para 2016 nos Estados Unidos é um valor muito próximo do milhão e setecentos mil. Um dos cancros que mais afeta as mulheres é o cancro da mama (Siegel *et al.*, 2016). Nos Estados Unidos em 2016 está previsto o aparecimento de 249.260 novos casos e que morram devido a este tipo de cancro 40.890 pessoas no decorrer do ano. Já em Portugal, são detetados anualmente cerca de 6.000 novos casos de cancro da mama, e 1.500 mulheres morrem com esta doença.

2.2 O cancro da Mama

A mama é uma glândula modificada da pele. Ela desenvolve-se na parede torácica, entre a clavícula e a sexta e oitava costelas. (Kopans, 2000). De acordo com *Center for Disease Control and Prevention*², caso não tenhamos em consideração alguns tipos de cancro da pele, o cancro da Mama nos Estados Unidos é:

- O cancro mais comum nas mulheres, não importa a raça ou etnia.
- O cancro que mais leva à morte as mulheres hispânicas.
- A segunda maior causa de morte por cancro entre mulheres de raça Branca, Negra, Asiática, Ilhas do Pacífico, Índio Americanas e nativas do Alasca.

Em 2012, ano em que os dados mais recentes estão disponíveis, 224.147 mulheres e 2.125 homens foram diagnosticados com cancro da mama nos Estados Unidos. A

²<http://www.cdc.gov/cancer/breast/statistics/>

estes números pode-se acrescentar que no ano de 2012, 41.150 mulheres e 405 homens morreram devido ao cancro da mama. Segundo Ferreira (2010) a partir de 1995 começou a verificar-se uma ligeira tendência para a diminuição da mortalidade devido à introdução de programas de rastreio, ou seja, a realização periódica de exames num grupo populacional assintomático com o principal intuito de detetar o cancro da mama num estado precoce, sendo que a mamografia é o exame imagiológico mais utilizado neste tipo de prevenção. De salientar que de acordo com Ferreira (2010) o estudo mais antigo sobre programas de rastreio está datado entre 1963 e 1970, ora, desde essa data até aos dias de hoje conclui-se que houve uma redução da mortalidade como consequência do cancro da mama na ordem dos 30% quando comparado com o grupo de controlo após cinco anos. É de acrescentar que nenhuma outra medida diagnóstica ou terapêutica levou a uma redução tão acentuada da taxa de mortalidade.

2.2.1 Mamografia

Nesta sub-secção queremos aprofundar um pouco o conceito Mamografia. Ora segundo Kopans (2000) o objetivo final da Mamografia por raios X é produzir imagens detalhadas das estruturas internas da mama, com a finalidade de permitir a deteção precoce do cancro da mama. Como é necessário um detalhe fino, o processo requer imagens com elevada resolução espacial. Sabendo-se que as diferenças inerentes de atenuação de raios X (contraste tecidual) entre tecidos mamários normais e doentes são muito pequenas, a mamografia de alta qualidade requer a capacidade de realçar tais diferenças e fornecer uma resolução de alto contraste (Kopans, 2000). Produzir estas imagens envolve uma iteração complexa de muitos fatores relacionados que fogem ao âmbito desta dissertação, mas que estão claramente explicados em Kopans (2000).

Nos rastreios descritos anteriormente, de acordo com Ferreira (2010), é de salientar que nos que obtiveram melhores resultados existiu uma influência de forma inequívoca da qualidade da mamografia e da experiência dos radiologistas. A *American Cancer Society* recomenda a realização de um exame clínico e de uma mamografia de dois em dois anos em pessoas que se encontrem na faixa etária dos 40 aos 49 anos. A partir dos 50 anos, de acordo com a mesma instituição, estes exames deverão ser realizados com periodicidade anual. De acordo com Orvalho (2008) a mamografia anual e o exame clínico podem detetar simultaneamente aproximadamente 80% de cancros da mama. De alertar que apesar da taxa de deteção alta, os 20% que a mamografia anual e o

exame clínico não detetam tornam-se palpáveis no ano seguinte, representando cancros de intervalo, que tendem a ser agressivos.

De acordo com Dores (2014) a mamografia mostra as diferenças nas densidades teciduais da mama. Ora estas diferenças fornecem a base radiográfica da mama. Os tecidos glandulares e fibrosos (ou conjuntivos), tecidos estes que são mais densos, aparecem como estruturas ou regiões com cor clara. Por sua vez, o tecido adiposo (gordura), tecido este que é menos denso, aparece em tons que variam entre o cinza-claro e o cinza-escuro, dependendo da sua espessura. A Figura 2.1 ilustra esta diferença tecidual.

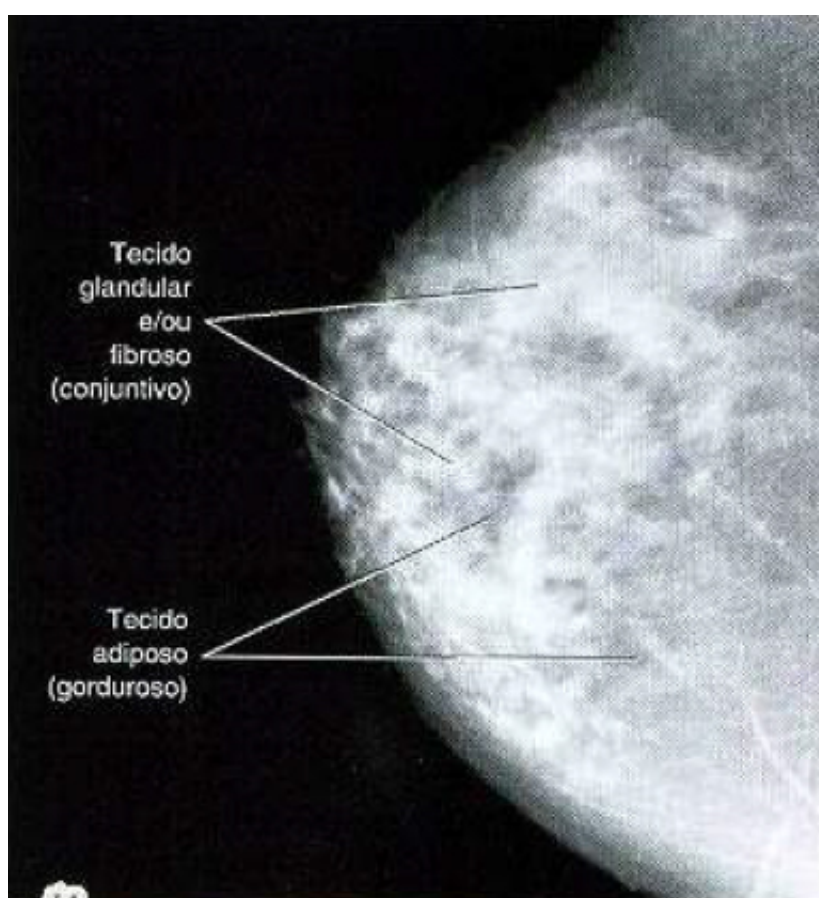


Figura 2.1: Diferenças entre tecidos mamários presentes numa mamografia. Retirado de Bontrager (2006)

As mamas descrevem-se como densas ou adiposas, segundo predominância de tecido fibroconjuntivo ou tecido gordo, podendo descrever-se tanto a morfologia, o volume e a situação que os mesmos ocupam nas mamas, sofrendo estas uma modificação segundo

a idade (Pisco, 2001). A Figura 2.2 ilustra a evolução da mama de acordo com a perspetiva mamográfica.

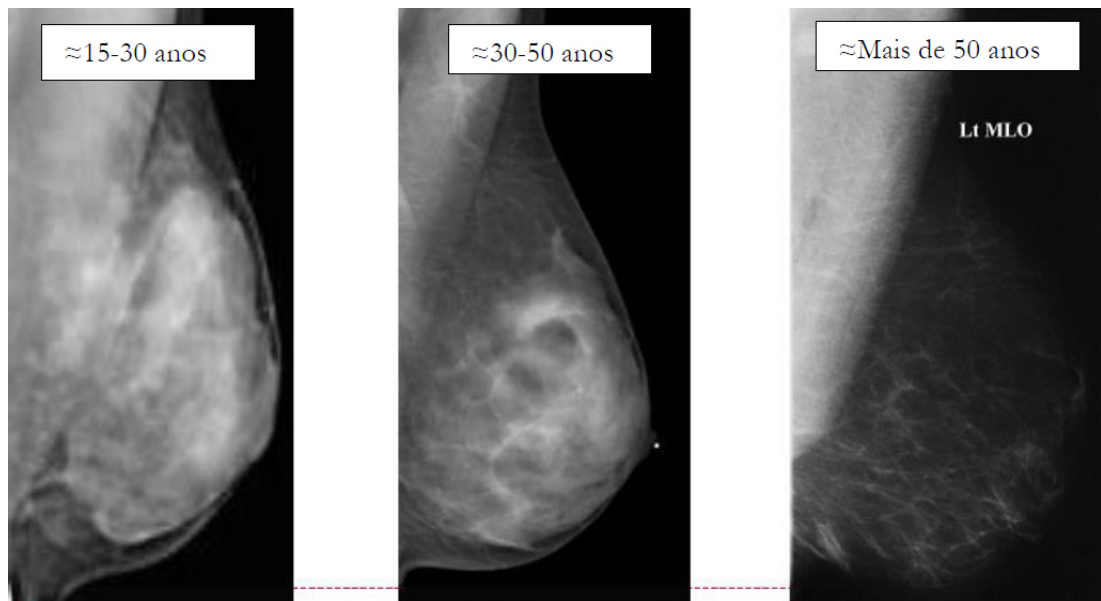


Figura 2.2: Evolução da mama segundo a idade. Retirado de Dores (2014)

Como visto anteriormente a mamografia pode ajudar um determinado médico a detetar tumores quando estes ainda não podem ser detetados por palpação. Ora, tal só é possível devido à precisão da mamografia. Esta precisão permite ao médico conhecer o tamanho, localização e características de nódulos com apenas alguns milímetros. O sistema BI-RADS[®] (*Breast Imaging Reporting and Data System*) foi publicado pelo *American College of Radiology* (ACR) em 1993³ e é um sistema que visa padronizar os relatórios médicos relativos à mamografia, permitindo desta forma diminuir as discordâncias existentes na interpretação de mamografias. O sistema BI-RADS[®] disponibiliza uma série de descritores que caracterizam os "achados" observados. Este sistema permite ainda definir categorias que resumem as principais conclusões a que o radiologista, ao avaliar o exame médico, pode chegar.

2.2.2 Sistema BI-RADS[®]

O sistema BI-RADS[®] é um sistema de classificação que foi desenvolvido para promover uma uniformização dos relatórios mamográficos. Ora segundo Cunha (2011) a falta

³Em 1995, 1998, 2003 e 2013 foram publicadas novas edições do sistema BI-RADS[®]

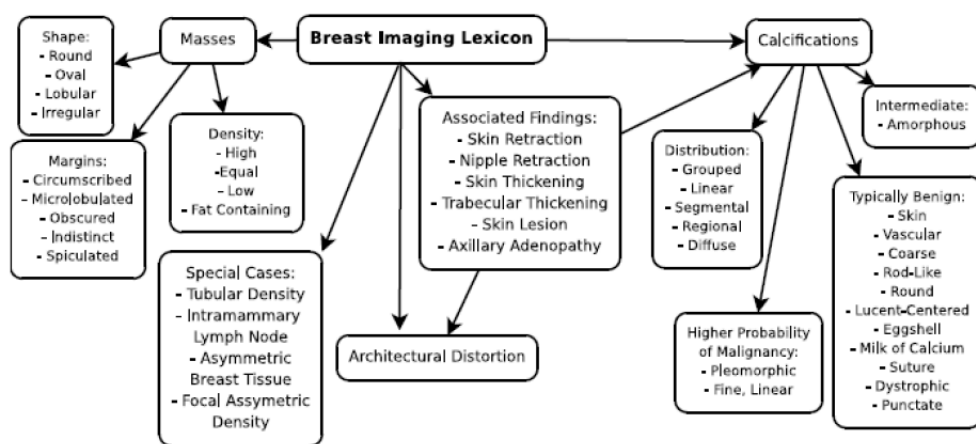


Figura 2.3: Lévico BI-RADS[®] adaptado de Nassif *et al.* (2009)

Round: Redonda	Massa esférica, em forma de bola, circular ou de forma globular.
Oval: Oval	Massa elíptica ou em forma de ovo.
Lobular: Lobular	Massa que possui contornos com ondulações.
Irregular: Irregular	A forma da lesão não pode ser caracterizada por nenhuma das anteriores.

Tabela 2.1: Lévico BI-RADS Forma, adaptado de Cunha (2011)

desta uniformização resulta muitas vezes em relatórios ambíguos que podem interferir na estratégia de conduta e tornar o controlo evolutivo difícil, surgindo com isso, dificuldades na interpretação de mamografias.

O léxico BI-RADS[®] é constituído por 43 descritores, ilustrados na Figura 2.3. Estes descritores são por sua vez organizados numa hierarquia.

No trabalho realizado por Cunha (2011) uma técnica de radiologia, Inês Moreira, traduziu cada um dos descritores BI-RADS[®] para a língua portuguesa que resultou na tradução apresentada nas tabelas seguintes, tabelas estas onde constam cada um dos descritores, *Shape*, *Margins*, *Density*, *Calc. Morphology*, *Calc. Distributions*, *Special Cases* e *Associated Findings*, bem como os respetivos termos equivalentes em Português. Começemos pela massa, de acordo com Cunha (2011) a massa é uma lesão que ocupa um espaço com bordos externos convexos observada em pelo menos duas projeções. Segundo o léxico BI-RADS[®] a massa pode ser representada por Forma, Margem e Densidade, descritas nas tabelas 2.1, 2.2 e 2.3, respetivamente.

Circumscribed (Well-Defined or Sharply-Defined) Margins: Circunscritas	As margens estão bem demarcadas, com uma transição abrupta entre a lesão e o tecido circundante.
Microlobulated Margins: Microlobuladas	Margens onduladas com ciclos curtos produzindo pequenas ondulações.
Obscured Margins: Obscuras	Está oculta por sobreposto ou adjacente a tecido normal e não pode ser avaliada.
Indistinct (Ill Defined) Margins: Indistintas	Má definição das margens suscitam preocupações de que pode haver infiltração da lesão.
Spiculated Margins: Espiculadas	Lesão caracterizada por linhas que se irradiam a partir das margens de uma massa.

Tabela 2.2: Léxico BI-RADS Margens, adaptado de Cunha (2011)

High density Alta Densidade
Low density (lower attenuation, but not fat containing): Baixa Densidade
Equal density (isodense): Isodenso
Fat containing / radiolucent: Radioluciente

Tabela 2.3: Léxico BI-RADS Densidade, adaptado de Cunha (2011)

As calcificações são partículas de sais de cálcio que se depositam na mama devido ao leite ou secreção acumulada, infeção ou alguma inflamação no tecido mamário Cunha (2011). Ora, de acordo com a Figura 2.3, as calcificações podem ter diferentes tipos de distribuição. De acrescentar que as calcificações podem ser benignas, intermédias e com alta probabilidade de malignidade. Ora, segundo Cunha (2011) as calcificações benignas são geralmente maiores que as calcificações malignas. As calcificações malignas são geralmente pequenas e muitas vezes requerem o uso de uma lupa para serem vistas. A tabela 2.4 caracteriza as calcificações tipicamente benignas. Por sua vez, a tabela 2.5 caracteriza as calcificações que apresentam probabilidade de malignidade intermédia. A tabela 2.6 caracteriza as calcificações que apresentam uma alta probabilidade de malignidade.

No que concerne à distribuição das calcificações a tabela 2.7 ilustra os diversos tipos de distribuição. Já a tabela 2.8 caracteriza a possível presença de uma distorção arquitetural.

Skin: Calcificações Dérmicas	Pequenas, esféricas com centro radiotransparente. Mantêm uma relação fixa entre si nas diferentes mamografias.
Vascular Calcifications: Calcificações Vasculares	Calcificações lineares tubulares que estão claramente associadas com vasos sanguíneos.
Coarse or ("Popcorn Like"): Calcificações Grosseiras / Popcorn (pipoca)	São as calcificações clássicas produzidas por um fibroadenoma involutivo.
Large Rod-Like: Em forma de bacilo / bastonete	Normalmente benignas, formam barras contínuas que podem ser ocasionalmente ramificadas, possuem normalmente mais de 1mm de diâmetro.
Round: Calcificações Redondas	Quando múltiplas podem variar de tamanho. São geralmente consideradas benignas.
Lucent-Centered: Calcificações com centro lucente	Benignas. Têm superfícies lisas, redondas e ovais.
Eggshell or Rim: Calcificações em casca de ovo	Calcificações finas benignas que aparecem como cálcio depositado sobre a superfície de uma esfera.
Milk of Calcium: Calcificações Ductais ou leite cálcio	Microcalcificações sedimentadas dentro de pequenos quistos.
Suture: Calcificações Cicatriciais	São representadas por cálcio depositado
Dystrophic: Calcificações Distróficas	Forma irregular, geralmente têm mais de 0,5mm de tamanho. Têm normalmente centro lucente
Punctate: Calcificações Punctiformes	Normalmente menor que 0,5mm e possuem margens bem definidas.

Tabela 2.4: Léxico BI-RADS Morfologia (Tipicamente Benignas), adaptado de Cunha (2011)

Amorphous or Indistinct: Calcificações Amórficas ou Indistintas	São normalmente calcificações pequenas e de forma arredondada.
--	--

Tabela 2.5: Léxico BI-RADS Morfologia (Probabilidade de Malignidade Intermédia), adaptado de Cunha (2011)

Pleomorphic or Heterogeneous: Pleomórficas ou Heterogéneas	Normalmente mais visíveis que as amórficas, possuem tamanhos e formas variados que são geralmente inferiores a 0,5mm de diâmetro.
Fine, Liner or Fine, Linear, Branching: Calcificações Finas, Lineares	Calcificações irregulares curvilíneas ou lineares ou finas, que poderão ser descontínuas e inferiores a 0,5mm de diâmetro.

Tabela 2.6: Léxico BI-RADS Morfologia (Alta Probabilidade de Malignidade), adaptado de Cunha (2011)

Grouped or Clustered: Agrupada	Calcificações múltiplas que ocupam um volume pequeno (menos de 2cc) do tecido.
Linear: Linear	Calcificações alinhadas que podem ter pontos de ramificação. Esta distribuição é suspeita de elevada malignidade.
Segmental: Segmentar	Distribuição preocupante, visto que sugere depósitos no ducto e nos seus ramos, levantando a possibilidade de cancro da mama multifocal num lobo ou segmento de mama.
Regional: Regional	Calcificações dispersas num grande volume de tecido mamário.
Diffuse/Scattered: Difusa	Calcificações distribuídas aleatoriamente por toda a mama.

Tabela 2.7: Léxico BI-RADS Distribuição das Calcificações, adaptado de Cunha (2011)

Architectural Distortion : Distorção Arquitectural	Pode ser associada a uma massa, à assimetria ou calcificações. Na ausência de história adequada de trauma ou cirurgia, a distorção arquitectural é suspeita de malignidade e a realização de biopsia é apropriado.
---	--

Tabela 2.8: Léxico BI-RADS Distorção Arquitectural, adaptado de Cunha (2011)

Tubular Density/Solitary Dilated Duct: Dilatação Ductal	Habitualmente retroareolar; benigno, a menos que se associe a outros achados suspeitos.
Intramammary Lymph Node: Gânglio Intramamário	Lesão circunscrita com conteúdo adiposo.
Asymmetric Breast Tissue: Densidade Assimétrica	Assimetria do tecido mamário é julgada em comparação à área correspondente na outra mama e uma maior densidade do tecido mamário. Não há formação de massa focal, não tem densidade central, sem calcificações associadas.
Focal Asymmetric Density: Densidade Focal	Densidade que não pode ser descrita com precisão usando outras formas. É visível que possui um tecido com densidade assimétrica.

Tabela 2.9: Léxico BI-RADS Casos Especiais, adaptado de Cunha (2011)

A tabela 2.9 ilustra os termos usados no léxico BI-RADS[®] para caracterizar a presença de casos especiais.

Por último, a tabela 2.10 representa termos que podem ser usados com massas, assimetrias ou calcificações, ou então pode ser um achado isolado quando nenhuma anormalidade está presente.

No que concerne às categorias BI-RADS[®], estas dividem-se em sete tipos. Cada uma destas categorias sintetiza a opinião do radiologista em relação à lesão encontrada. Ora, assim sendo, os exames realizados pelos pacientes são classificados com base no grau de suspeita de lesões. A título de exemplo, a categoria 0 é atribuída quando o especialista acha que são necessários exames adicionais. Já a categoria 5 é atribuída a exames que sejam provavelmente malignos. A tabela 2.11 ilustra todas as categorias BI-RADS[®] existentes bem como a sua interpretação.

Segundo Cunha (2011) caso seja detetada uma massa suspeita, isto é, que possa ser cancerosa, o paciente poderá ser sujeito a uma biopsia. A biopsia trata-se de um procedimento (que poderá ser cirúrgico ou não) em que é recolhida uma amostra do nódulo suspeito. O tecido retirado é posteriormente analisado por um patologista com o objetivo de confirmar se a origem do nódulo em causa é de natureza maligna (Ferreira, 2010). A Figura 2.4 ilustra duas situações, a da esquerda onde é um exemplo de uma mamografia onde não é necessário realizar uma biopsia e a direita onde está

Skin Retraction: Retracção Cutânea	Retracção anormal da pele
Nipple Retraction: Retracção do Mamilo	Mamilo pode-se apresentar repuxado ou invertido. Quando aparenta ser crónico e na ausência de qualquer outro achado associado, não é sinal de malignidade.
Skin Thickening: espessamento cutâneo	Pode ser focal ou difusa e normalmente maior que 2mm.
Trabecular Thickening: Espessamento Trabecular	Espessamento dos septos ¹⁰ fibrosos da mama.
Skin Lesion: Lesão Cutânea	Pode ser confundida com uma lesão intramamária.
Axillary Adenopathy: Adenopatia/ Adenomegalia Axilar	Achado associado que pode ser usado em conjunto com outro achado para indicar que a estrutura de um tecido normal encontra-se distorcido ou com retracção.

Tabela 2.10: Léxico BI-RADS Achados Associados, adaptado de Cunha (2011)

Categoria	Interpretação
BI-RADS 0	Inconclusivo Necessário exame adicional
BI-RADS 1	Benigno Realizar controlo anual a partir dos 40 anos.
BI-RADS 2	Benigno - Realizar controlo anual a partir dos 40 anos.
BI-RADS 3	Provavelmente Benigno Realizar novo exame em seis meses (eventualmente biopsia).
BI-RADS 4	Suspeito biopsia.
BI-RADS 5	Provavelmente Maligno biopsia.
BI-RADS 6	Maligno Comprovado por biopsia.

Tabela 2.11: Categorias BI-RADS[®], adaptado de Cunha (2011)

representada uma mamografia onde é necessária a realização de uma biópsia para determinar qual a natureza do nódulo presente na mamografia.

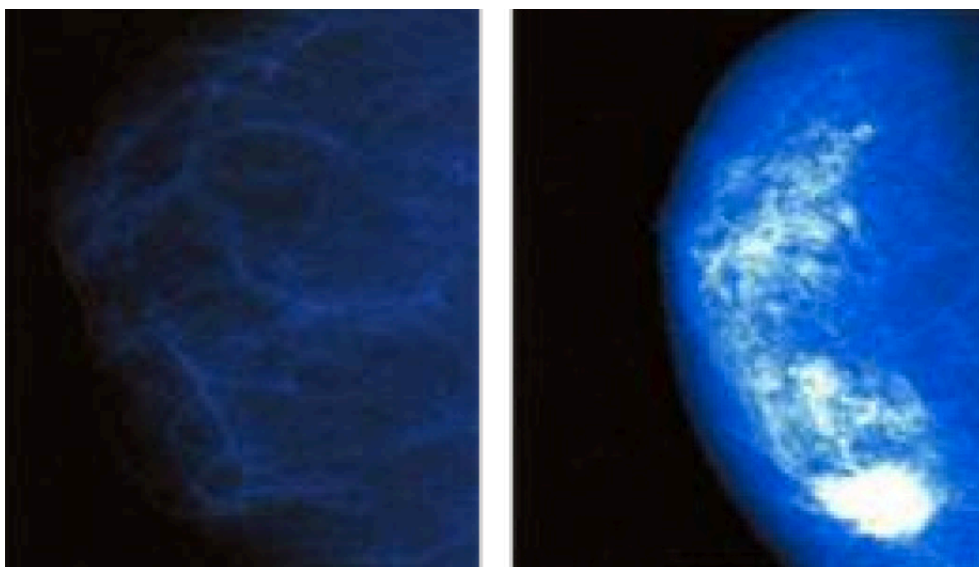


Figura 2.4: Imagens referentes a duas mamografias distintas. A mamografia da esquerda apresenta uma mama normal, em que as áreas mais densas (brancas) são os canais galactóforos. A mamografia da direita, por sua vez, apresenta uma área branca densa (canto inferior direito da imagem) que indica a presença de um tumor, retirado de Ferreira (2010)

Segundo Orvalho (2008) desde o início dos rastreios até aos dias de hoje, a percentagem de carcinoma ductal *in situ* (a fase mais precoce do cancro da mama) aumentou de 5%, para valores que se situam entre os 20% e os 30% de todos os cancros detetados. Posto isto, segundo Ferreira (2010) é fundamental conhecer alguns dos termos mais utilizados na descrição de tumores mamários, uma vez que o tratamento e prognóstico variam de doente para doente e em função do tipo de cancro. A Figura 2.5 representa a anatomia de uma mama normal. Ora, nesta imagem podemos visualizar a representação do ducto e do lóbulo da mama. Estes dois tecidos glandulares são importantes pois segundo Ferreira (2010) quase todos os tumores malignos da mama têm a sua origem nestes tecidos. Assim sendo, é possível constatar que os tipos mais frequentes de cancro da mama são o carcinoma ductal e o carcinoma lobular. De seguida é apresentada a definição presente em Ferreira (2010) sobre os diferentes tipos de tumores mamários, bem como, os respetivos termos científicos.

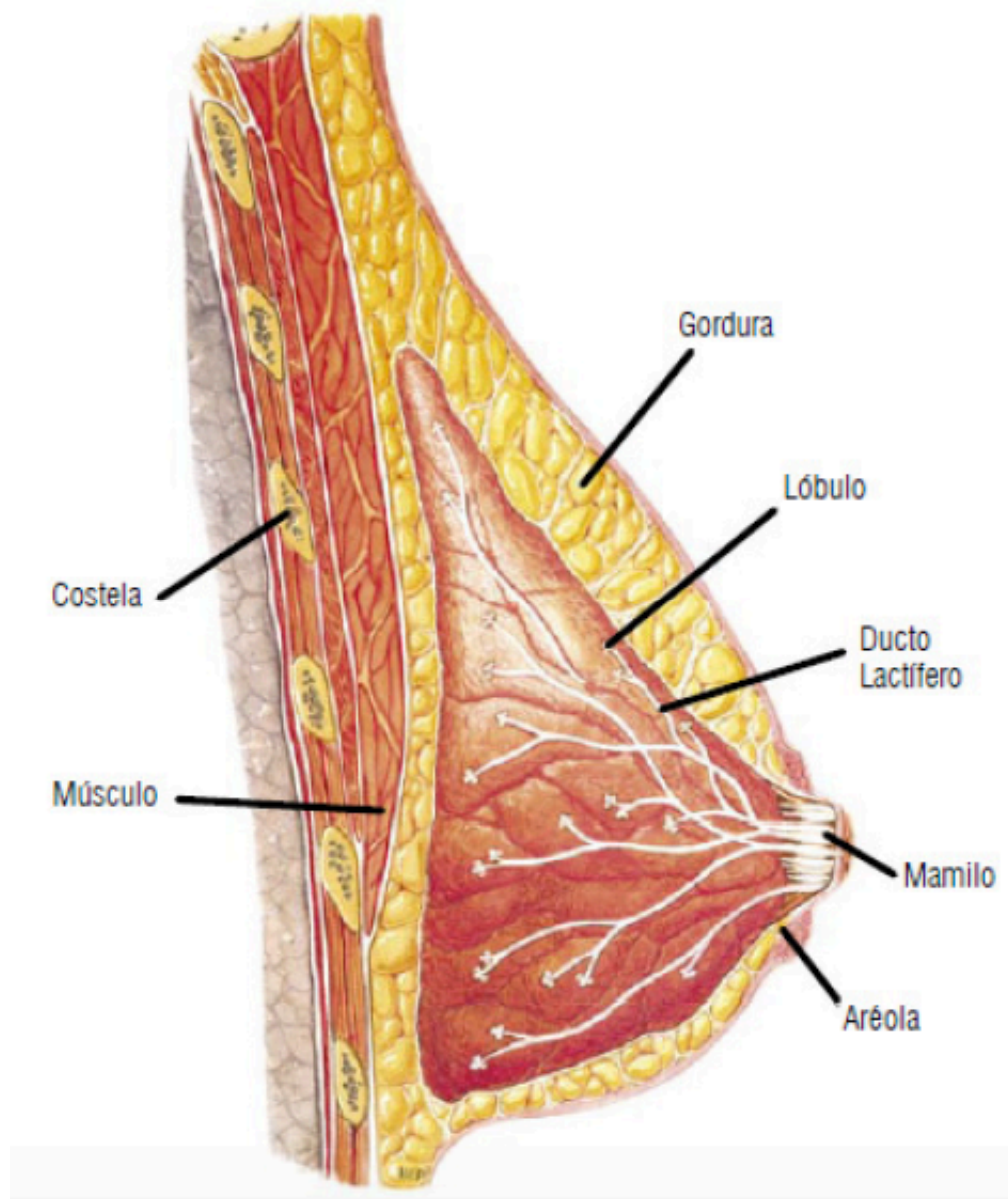


Figura 2.5: Anatomia de uma mama saudável, retirado de Ferreira (2010)

- ***In situ*** - Trata-se do cancro de mama precoce que se encontra limitado aos ductos (carcinoma ductal *in situ*) ou lóbulos (carcinoma lobular *in situ*), sem invasão dos tecidos mamários vizinhos e de outros órgãos.
- **Carcinoma ductal *in situ* (CDIS)** - Trata-se do cancro da mama não invasivo mais frequente. Praticamente todas as mulheres com CDIS têm hipóteses de cura. Ora, a mamografia é nesta fase precoce o melhor método para diagnosticar o cancro.
- **Carcinoma lobular *in situ* (CLIS)** - Embora não seja verdadeiramente cancro, o CLIS é de forma habitual classificado como um cancro da mama não invasivo. De salientar que diversos especialistas defendem que o CLIS não se transforma num carcinoma invasor. Contudo, as mulheres com esta neoplasia apresentam um risco maior de desenvolver cancro da mama invasor.
- **Carcinoma ductal invasor (CDI)** - Trata-se do cancro da mama invasor que se verifica com mais frequência. Tem origem nos ductos e invade tecidos vizinhos. Nesta fase pode disseminar-se através do sangue ou dos vasos linfáticos, acabando por atingir outros órgãos. Cerca de 80% dos cancros invasores são carcinomas ductais.
- **Carcinoma lobular invasor (CLI)** - Este tipo de cancro tem origem nas unidades produtoras de leite, ou seja, nos lóbulos. À semelhança do CDI, este pode disseminar-se (metastizar) para outras partes do corpo. Cerca de 10% dos cancros de mama invasores são carcinomas lobulares.
- **Carcinoma inflamatório da mama** - Trata-se de um cancro extremamente agressivo, contudo, pouco frequente. Este tipo de cancro corresponde a cerca de 1% a 3% de todos os cancros de mama.

De salientar que existem outros tipos de cancro, como por exemplo o Carcinoma Medular, o Carcinoma Mucinoso, o Carcinoma Tubular, o Tumor Filóide Maligno, entre outros, que são cancros da mama mais raros que os descritos acima.

2.3 Arquitetura Cliente-Servidor

Nesta secção será apresentada de forma breve a arquitetura Cliente-Servidor. Achamos que esta breve explicação contribui com conceitos que são necessários para melhor entendimento desta dissertação.

In the context of a communication session between a pair of processes, the process that initiates the communication (that is, initially contacts the other process at the beginning of the session) is labeled as the client. The process that waits to be contacted to begin the session is the server (Kurose & Ross, 2005).

Segundo Coulouris *et al.* (2011) a Arquitetura Cliente-Servidor é uma arquitetura onde clientes requerem e recebem serviços de um servidor centralizado. O servidor pode ser cliente de outro servidor no caso de requerer informação de outro servidor. Os computadores que fazem o papel de clientes fornecem ao seu utilizador uma interface que permite requisitar serviços ao servidor e mostrar os resultados que o servidor retorna. Já o papel do servidor é aguardar por pedidos de clientes e depois responder aos mesmos. Idealmente o servidor fornece uma interface transparente e standardizada aos clientes para que estes não se tenham que preocupar com as especificidades do sistema que está a fornecer o serviço, ou seja, sem que os clientes tenham que se preocupar qual o *software* e qual o *hardware* que constitui o servidor. Neste tipo de arquitetura o cliente é muitas vezes uma *workstation* ou um computador pessoal, enquanto que os servidores estão alocados em algum lado da rede, sendo usualmente uma ou mais máquinas com elevados poderes computacionais. Esta arquitetura é especialmente eficaz quando quer o cliente quer o servidor executam tarefas distintas. Por exemplo, no processamento de dados num hospital, o computador cliente pode estar a executar uma aplicação cuja finalidade é inserir informação dos pacientes, enquanto isso, o computador servidor está a executar uma aplicação que faz a gestão da base de dados em que a informação dos pacientes é permanentemente guardada. Muitos clientes podem aceder à informação do servidor simultaneamente, e além disso cada um destes clientes pode estar a executar outras tarefas como por exemplo a tarefa de enviar um *e-mail*.

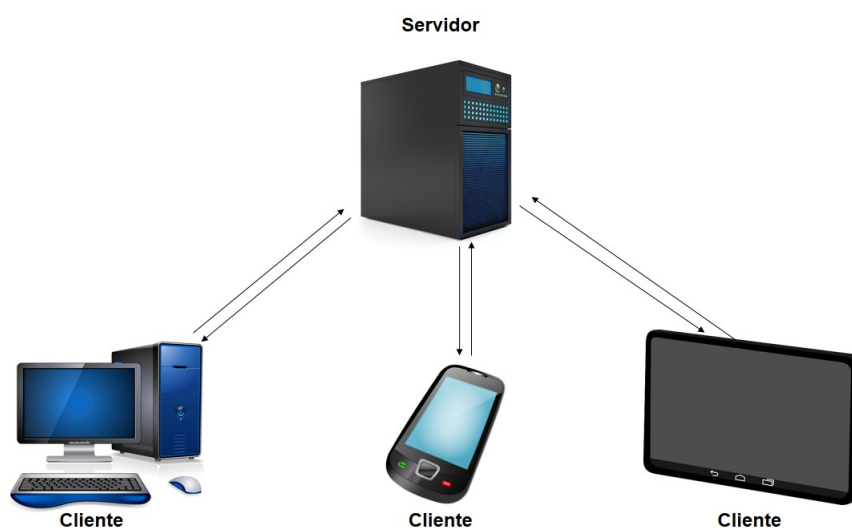


Figura 2.6: Representação de uma arquitetura Cliente-Servidor

No caso da Web segundo Kurose & Ross (2005), o *browser* inicializa o contacto com o servidor Web, sendo assim, o *browser* é o cliente, enquanto que o servidor Web é o servidor. A Figura 2.6 ilustra a representação de uma Arquitetura Cliente-Servidor.

2.4 *Speech-To-Text*

Antes de iniciarmos a explicação do que é uma ferramenta *Speech-To-Text* queremos salientar que embora sejam muitas vezes confundidos na literatura, o termo *Speech Recognition* (reconhecimento do discurso) significa algo diferente que o termo *Voice Recognition* (reconhecimento de voz). De acordo com Reynolds *et al.* (2000) o reconhecimento de voz está relacionado com reconhecer e identificar de forma individual a voz de uma pessoa, ou seja, reconhecer se é a pessoa A que é o locutor. Este tipo de reconhecimento não é dependente do idioma. Por outro lado, o reconhecimento de discurso é dependente do idioma e não tem como objetivo reconhecer quem é o locutor, mas sim, detetar palavras no conteúdo do discurso. O trabalho elaborado nesta dissertação tem como objetivo extrair palavras relevantes provenientes de textos ditados pelo que estamos interessados em ferramentas de *Speech Recognition*.

As ferramentas de *Speech-To-Text* permitem a conversão de discurso falado em texto escrito. De acordo com Rabiner & Juang (1993) estas ferramentas requerem conhe-

imento de um grande número de temáticas, um número tão elevado que dificilmente pode ser do conhecimento de apenas uma pessoa. Para realizar tal processo, este tipo de ferramentas necessitam de estar dotadas de conhecimento na área da linguística, na área de ciência de computadores e na área de engenharia eletrotécnica, entre outras.

As ferramentas de *Speech-To-Text* compreendem diversas etapas para poderem funcionar, sendo que cada uma delas é responsável por uma função diferente. A primeira etapa é o processamento de sinal. De acordo com Gold *et al.* (2011) nesta etapa existe uma retificação das ondas sonoras recebidas de forma a reduzir os efeitos da captura do som, ou seja, o canal do áudio, bem como, o ruído, que inclui o ruído produzido pelo microfone no processo de captura de áudio e o ruído resultante do ambiente em redor do utilizador. Nesta fase existe ainda uma parametrização da forma das ondas sonoras de forma a modelar e extrair informação relevante, facilitando a sua análise nas etapas seguintes. Apesar de existirem técnicas analógicas de captura de som, nesta etapa é necessário que o som seja convertido para o formato digital, dividindo o *input* em segmentos distintos, suficientemente curtos para minimizar o efeito da conversão.

A etapa seguinte neste processo é o modelamento acústico. Segundo Huang *et al.* (2001) nesta etapa recorre-se a modelos probabilísticos para calcular a possibilidade de um determinado sinal corresponder a uma sonoridade. Como analisado na etapa anterior o sinal obtido da voz transporta um elevado número de características que são bastante variáveis. Assim sendo, a tarefa de estimação de probabilidade revela-se bastante complexa. Em termos de solução, nos primórdios do estudo do reconhecimento de discurso, a mais usada era baseada em algoritmos *Dynamic Time Warping* (DTW). Atualmente o modelo acústico que evidencia maior eficiência é baseado em *Hidden Markov Models* (HMM) (Huang *et al.* , 2001). De salientar que a *Google* e a *Nuance* utilizam *Neural Networks* nesta etapa.

A etapa seguinte é a utilização de um modelo linguístico. De acordo com Adami (2010) esta etapa consiste no recurso a uma sequência pré-determinada de palavras que ocorrem regularmente no discurso para aumentar as hipóteses de acerto. Como exemplo, suponhamos que existe duas ou mais palavras que ocorrem sempre seguidas uma da outra, assim sendo as hipóteses de dedução do programa resumem-se a aparecerem as duas palavras ou nenhuma delas.

Como etapa final, temos o decodificador. Esta componente unifica o modelo acústico e o modelo linguístico. Este programa usa as estatísticas obtidas através das duas com-

ponentes para determinar quais são as palavras que correspondem mais provavelmente aos dados acústicos fornecidos.

A forma como funciona internamente uma ferramenta de *Speech-To-Text* é bastante complexa, pelo que, a descrição feita nesta dissertação serve como uma explicação básica de como o processo funciona. Os leitores interessados numa explicação detalhada podem consultar o trabalho apresentado por Adami (2010).

De salientar que o uso de ferramentas de *Speech-To-Text* é extremamente importante para pessoas que necessitem de gerar conteúdo em formato textual mas que não queiram fazer inserção manual dos dados, sendo também muito importante e útil para pessoas com limitações que as impeçam de usar um teclado.

Capítulo 3

Estado da Arte e Trabalhos Relacionados

A tecnologia de reconhecimento de discurso tem sido melhorada ao longo dos anos, permitindo relatórios precisos e conseqüentemente o armazenamento de mais informação de teor qualitativo. Este capítulo visa apresentar alguns trabalhos que envolvem o uso de tecnologias de reconhecimento de discurso aplicadas à área médica.

Existem trabalhos na literatura que reportam histórias de sucesso provenientes do uso de reconhecimento de discurso com o objetivo de extrair o conteúdo de textos ditados. Como exemplo temos o trabalho apresentado por Kang *et al.* (2010) em que usam a tecnologia de reconhecimento de discurso na área da patologia cirúrgica e concluem que esta tecnologia é útil no fluxo de trabalho da anatomia patológica. Kang *et al.* (2010) concluem ainda que a tecnologia de reconhecimento de discurso representa um bom retorno do investimento, uma redução nos erros, bem como, uma poupança de custos. O trabalho de Kang *et al.* (2010) reporta a experiência de utilização da tecnologia de reconhecimento de discurso na área de patologia cirúrgica de um grande centro multi-institucional. Como ferramenta *Speech-To-Text* usam o *Dragon*, ferramenta esta que irá ser apresentada no decorrer desta dissertação. Neste centro médico cada utilizador da ferramenta tem que passar por um processo de treino de uma hora e por um processo de criação de um perfil de voz que dura cerca de dez minutos. Nesta ferramenta, o utilizador pode preencher todo o relatório com a voz. Os autores chegam ainda à conclusão que ao usar ferramentas de reconhecimento de discurso os erros nos relatórios diminuem. Já como desvantagem, surge o caso em que

o utilizador possa hesitar ou fazer pausas no discurso o que pode levar o *software* a levar essas pausas ou hesitações à letra, levando assim à produção de relatórios em que o conteúdo não é o desejado.

Apesar de existirem casos de sucesso, os sistemas de reconhecimento de discurso apenas funcionam bem quando o vocabulário é limitado e os ditados são realizados num ambiente controlado, tal como acontece nas áreas de radiologia ou de patologia (Henricks *et al.* , 2002). Neste estudo foi avaliada a capacidade de um sistema de reconhecimento de voz na introdução de descritores de biopsias. Ora, neste sistema, cada utilizador tem que participar numa sessão de treino que dura entre duas a três horas. O foco deste sistema é o preenchimento de *templates* de relatórios, sendo que 99% da informação inserida ao longo do estudo pertence a um *template* enquanto que os restantes 1% correspondem a texto livre. Através deste estudo Henricks *et al.* (2002) constataram que ao fim de 18 meses de utilização da ferramenta de reconhecimento de discurso, 70% das descrições foram realizadas através de *Speech-To-Text*, enquanto que apenas 30% foram feitas de forma manual. De salientar que a *accuracy* do *Speech-To-Text* está entre os 70% e os 90%.

Yen & Bakken (2012) salientam que o reconhecimento de discurso é bastante desadequado para usar em ambientes públicos com elevado nível de barulho, pois neste tipo de ambientes, além da sua performance ser baixa, a confidencialidade da informação de saúde do paciente fica comprometida.

Alguns trabalhos na literatura não são a favor da utilização de tecnologia de reconhecimento de discurso na área da radiologia e conseqüentemente reportam uma taxa de erro elevada referente ao texto reconhecido do Toit *et al.* (2015); Basma *et al.* (2011); Hoyt & Yoshihashi (2010); McGurk *et al.* (2008); Hammana *et al.* (2015). Mais especificamente do Toit *et al.* (2015) testaram a *accuracy* do reconhecimento do discurso aplicado a relatórios de radiologia num hospital-escola multi-linguístico situado na África do Sul. Para tal fizeram uma comparação entre 300 relatórios obtidos através de reconhecimento de discurso com 300 relatórios provenientes da transcrição de um ditado por parte de uma pessoa. Posto isto, os resultados indicam que 25,6% dos relatórios obtidos por reconhecimento de discurso, ou seja, 77 relatórios, contêm erros, sendo que 9,6% (29 relatórios) contêm erros clinicamente significativos. Por sua vez, apenas 9,3% dos relatórios obtidos através da transcrição de um ditado, ou seja, 28 relatórios possuem erros, sendo que, apenas 2,3% (7 relatórios) contêm erros

cl clinicamente significativos. Foi realizado um estudo *follow-up* passados 4 anos em que 24,3% dos relatórios, ou seja, 73 relatórios obtidos por reconhecimento de discurso apresentavam erros, sendo que 6% destes relatórios possuem erros clinicamente significativos. Estas diferenças entre o estudo inicial e o *follow-up* não são estatisticamente significativas. É de acrescentar que o *software* em uso neste hospital requer que cada utilizador passe por uma seção prévia de treino. É também referido neste estudo que os utilizadores que tenham a língua inglesa como língua nativa têm melhores resultados do que utilizadores que tenham a língua inglesa como segunda língua.

Basma *et al.* (2011) realizaram uma comparação entre reconhecimento de discurso e transcrição de ditado aplicado à área de relatórios de imagem mamária. Neste estudo chegaram à conclusão que em 23% dos relatórios obtidos por reconhecimento de discurso existiu pelo menos um erro grave. Em contrapartida nos relatórios obtidos por transcrição de ditado apenas em 4% dos relatórios existiu pelo menos um erro grave. Neste estudo concluíram que os erros não diferem substancialmente entre os utilizadores que têm a língua inglesa como língua nativa e os utilizadores que tenham a língua inglesa como segunda língua. É de salientar que os relatórios usados neste estudo possuem um grau de complexidade mais elevado que os relatórios produzidos através de *template*. Em termos de *software* foi usado o *Speech Magic*, produzido pela *Nuance*. Para captura do áudio proveniente do utilizador foi usado um microfone profissional. O erro mais comum produzido pelo reconhecimento do discurso é a adição de uma palavra à frase. Os autores concluem assim que os relatórios de imagem mamária gerados através de reconhecimento de discurso apresentam 8 vezes mais probabilidade de erro que os relatórios obtidos através da transcrição de ditados.

Hoyt & Yoshihashi (2010) fizeram um estudo onde refletem sobre a implementação de um sistema de reconhecimento de discurso usado para documentar os registos de saúde de um hospital militar e das suas 12 clínicas periféricas. Os autores focam a importância de substituir o método tradicional das clínicas, que consiste em o médico ditar o relatório, um transcritor traduzir o relatório para texto e depois que o devolva ao médico para validação. Este método pode demorar várias dias a estar completo, o que não é o ideal quando o acesso rápido ao relatório é necessário. Nestas clínicas o *software* usado é o *Dragon NaturallySpeaking (version 9)*. Trata-se de uma versão médica que inclui 14 vocabulários pré-configurados para especificidades médicas. Após a implementação do serviço, 75 clínicos voluntariaram-se para usar a

ferramenta de reconhecimento de discurso sendo que, pós implementação, foi realizado um questionário onde participaram 85% (64 clínicos) com o intuito de estudar a quantidade de clínicos que deixaram de usar a ferramenta de reconhecimento de discurso, sendo que 69% (44 clínicos) continuaram a usar a ferramenta, tendo os restantes 31% (20 clínicos) deixado de usar a ferramenta. Foi possível concluir que os que continuaram a usar a ferramenta fizeram-no por acharem tratar-se de um *software* preciso, mais rápido do que digitar o texto, notaram uma melhoria da qualidade das anotações e que concluíram que conseguiam finalizar o encontro com o paciente no mesmo dia. Por seu turno, os clínicos que deixaram de usar a ferramenta fizeram-no devido a acharem que tiveram um treino inadequado, por acharem que diminuíram a produtividade devido à imprecisão do programa e alguns afirmaram que com o seu uso não melhoram a qualidade das anotações. De salientar que 70% dos clínicos que deixaram de usar o programa salientaram que o fizeram devido ao tempo necessário para corrigir os erros gerados pelo *software* de reconhecimento do discurso.

O objetivo do estudo de McGurk *et al.* (2008) era confirmar se os relatórios do departamento de radiologia gerados através de reconhecimento de voz, têm mais erros que os relatórios gerados através do método tradicional de ditado e transcrição. Neste estudo foram analisados todos os relatórios gerados durante uma semana num hospital britânico. Durante uma semana foram gerados 1887 relatórios, sendo que 61,5% (1160 relatórios) foram gerados através de reconhecimento de discurso. Por sua vez, 38,5% (727 relatórios) foram gerados através de ditado e transcrição do relatório. Analisando todos os relatórios foi possível constatar que 3,8% (71 relatórios) possuíam erros, sendo que 56 desses relatórios foram gerados através do reconhecimento de discurso. Este número significa que 4,8% dos relatórios gerados por reconhecimento de discurso possuem erro. No que concerne aos relatórios ditados e transcritos, em 15 deles foram identificados erros, o que significa que 2,1% dos relatórios ditados e transcritos possuem erros. A diferença de erros nos relatórios entre os dois métodos de geração é estatisticamente significativa ($p=0.002$). Dos 71 relatórios que possuem erros, 52,1% (37 relatórios) possuem erros que afetam o entendimento do relatório. Através deste estudo foram também identificados fatores que afetam a probabilidade de erros nos relatórios gerados através do reconhecimento de discurso, sendo um destes fatores o facto de trabalhar num ambiente ocupado ($p < 0.001$) e outro fator o facto de ter uma língua que não a inglesa como primeira língua ($p=0.034$). Por sua vez o grau do operador que realizou o ditado não teve significância no aumento

da quantidade de erros. Deste modo, os autores concluem que erros em relatórios gerados através de reconhecimento de discurso são mais prováveis que aconteçam em ambientes barulhentos, com elevada carga de trabalho e que são mais prováveis de acontecer quando o ditado é feito por radiologistas que não têm a língua inglesa como primeira língua.

Por sua vez Hammana *et al.* (2015) fizeram uma revisão sistemática a artigos relevantes sobre o impacto dos sistemas de reconhecimento de discurso nas taxas de erro dos relatórios, bem como na produtividade dos departamentos de radiologia. Em termos de período temporal, foram incluídos neste estudo artigos publicados entre o ano de 1992 e o ano de 2013. A partir destes relatórios os autores chegaram a algumas observações. A primeira delas prende-se com o facto de ser claro que uma certa porção de relatórios contém pelo menos um erro. Foi também possível observar que, dos relatórios que contém erros, a percentagem de erros que afeta o entendimento varia de 4% até 52%. A alta proporção de relatórios a conter pelo menos um erro leva a crer que os sistemas de reconhecimento de discurso são imperfeitos e que para produzir relatórios sem erros o radiologista necessita de perder tempo na correção dos erros diminuindo assim a produtividade individual. Foi também concluído pelos autores que apesar da produtividade individual do clínico diminuir com o uso de tecnologias de reconhecimento de discurso a produtividade do departamento em geral aumentou, pois os relatórios ficam disponíveis de forma mais rápida. De salientar que foi notado pelos autores que a taxa de erro dos relatórios parece variar consoante a complexidade do relatório.

Os estudos apresentados anteriormente salientam algumas lacunas presentes nos sistemas de reconhecimento de discurso, principalmente questões relacionadas com a quantidades de erros presentes num relatório que seja gerado por este sistema. Contudo deles também advêm aspetos positivos. Um deles é sem dúvida o tempo que demora até um relatório ficar disponível, quando comparado com o sistema de ditado e transcrição (método tradicional), os sistemas de reconhecimento de discurso permitem obter o relatório de forma bastante mais rápida. Patel *et al.* (2013) realizaram um estudo onde tentam diminuir a taxa de erro fazendo sucessivas revisões dos textos ditados. Os relatórios dos médicos que participaram na experiência eram avaliados por outros radiologistas que classificam os erros consoante a sua gravidade e marcavam os erros nos relatórios. A cada dois meses era dado a cada radiologista que participou na

experiência os seus relatórios com os erros marcados, bem como a sua taxa de erro e ainda sugestões de como ele poderia melhorar o seu ditado. Tais sugestões eram por exemplo, sobre o correto posicionamento do microfone. Foi concluído que após a primeira intervenção os *scores* de cada radiologista aumentaram, querendo isto dizer que os seus relatórios passaram a conter menos erros. A partir da primeira intervenção a taxa de erro estabilizou. Os autores concluem ainda que alguns padrões de erro tendem a ser únicos de cada radiologista.

Todos estes estudos apresentados focam-se no texto em si, e não em palavras que poderiam ser extraídas do texto, com o intuito de construir dados estruturados que possam ser usados em estudos automáticos posteriores. Como visto no capítulo anterior, para descrever achados numa mamografia o radiologista usa a terminologia BI-RADS®. Tais descritores são bastantes úteis uma vez que através desses dados podem ser construídos modelos preditivos de forma automática e assim ajudar o clínico a suportar decisões clínicas. Deste ponto de vista, e apesar dos relatórios ditados para sistemas de reconhecimento de discurso terem uma elevada taxa de erro, quando é considerado o texto como um todo, eles podem ajudar a acelerar o processo de criar dados estruturados, já que o foco é em palavras relevantes e não no texto como um todo.

MammoClass Ferreira (2010) é um sistema de apoio à decisão clínica que utiliza dados estruturados (características BI-RADS® anotadas de imagens de mamografias) para fazer a classificação de um achado imagiológico. Estes dados estruturados podem ser introduzidos no sistema através de um formulário. Para esta ferramenta, o texto em si não é relevante. É suficiente que a ferramenta consiga extrair os termos BI-RADS® importantes contidos no relatório. Como visto anteriormente neste trabalho, estendemos o MammoClass para dar suporte ao texto ditado e digitado.

Capítulo 4

Ferramentas de *Speech-to-Text*

O *software Speech-To-Text* como visto anteriormente é um *software* que recolhe áudio e transforma esse áudio em texto. Este tipo de *software* é extremamente valioso para alguém que necessite de gerar conteúdo escrito sem a preocupação de escrever manualmente o texto. De salientar que é também importante para pessoas que devido a algum tipo de deficiência ou incapacidade estejam impedidas de usar um teclado.

Para esta dissertação, fizemos uma análise a várias ferramentas de *Speech-To-Text* com o intuito de escolher a que melhor se adequava ao nosso trabalho, ou seja, a que cumprisse os requisitos de reconhecer as palavras Portuguesas, mais particularmente termos pertencentes ao léxico BI-RADS[®], com uma taxa de acerto razoável. Outro dos requisitos era a necessidade de a ferramenta a usar ser gratuita.

As várias ferramentas analisadas são apresentadas de seguida, de salientar que esta análise aconteceu em Janeiro de 2015 e foram pesquisadas as ferramentas com maior relevo a essa data.

- ***Free Voice to Text***¹ - Este é um *software* que permite o envio de emails apenas ditando o texto que pretendemos enviar. Tal como o nome indica estamos perante um *software* livre. Em termos de suporte de idiomas, este *software* suporta a língua Inglesa, Espanhola, Francesa e Japonesa.

¹http://download.cnet.com/Free-Voice-to-Text/3000-7239_4-76115951.html

- **Talking Desktop**² - Este é um *software* que para além da função de transcrever áudio para texto possui funções para ditar as horas bem como avisos meteorológicos. Como problemas, é de salientar os poucos controlos que possui e um tempo de resposta lento. Trata-se de um *software* proprietário e o seu preço é de 47\$. No que concerne ao suporte de idiomas, este *software* suporta a língua Inglesa, Espanhola, Francesa e Alemã.
- **Dragon Naturally Speaking Home/Premium**³ - Através da nossa análise é possível dizer-se que este *software* é bastante assertivo. O mesmo possui uma interface amigável do utilizador e bastante funcional. Trata-se de um *software* proprietário e o seu preço é de 99,99\$ para a versão *Home* e de 199,99\$ para a versão *Premium*. De salientar a existência de uma versão *Medical* cujo preço é de 999,99\$. No que respeita ao suporte de idiomas, este *software* suporta a língua Inglesa, Espanhola, Francesa, Italiana entre outras, contudo, nesta lista de linguagens suportadas não está incluída a língua Portuguesa.
- **Freesr Speech Recognition**⁴ - Este é um *software* que permite ditar múltiplos textos, ou seja, permite que seja atribuído um número a cada janela e depois permite ao utilizador ditar para cada uma delas. É um *software* proprietário e o seu preço apenas pode ser obtido a pedido, contudo possui uma versão de demonstração. Em termos de suporte de idiomas é um *software* que apenas suporta a língua Inglesa.
- **Simon**⁵ - Este *software* é gratuito e de código aberto. Está disponível para Windows e Linux. Em termos de suporte de idiomas este é também um *software* que apenas suporta a língua Inglesa.
- **Web Speech API**⁶ - Esta ferramenta é uma API desenvolvida pela Google que permite ao programador obter a tradução da voz para texto. Esta ferramenta é gratuita e pode ser usada em qualquer sistema operativo desde que este permita o uso do *browser Google Chrome*. No que concerne ao suporte de idiomas, esta API, suporta muitas línguas, sendo que uma delas é a língua Portuguesa.

²<http://voice-recognition-software-review.toptenreviews.com/talkingdesktop-review.html>

³<http://www.nuance.com/for-business/by-product/dragon/product-resources/edition-comparison/index.htm>

⁴<http://freesr.org/>

⁵<https://simon.kde.org/>

⁶<https://dvcs.w3.org/hg/speech-api/raw-file/tip/speechapi.html>

Software	Gratuito	Preço (US\$)	Língua	Plataforma
Free Voice to Text	Sim	0	Inglesa, Espanhola, Francesa e Japonesa.	Windows
Talking Desktop	Não	47	Inglesa, Espanhola, Francesa e Alemã.	Windows
Dragon Naturally Speaking Home	Não	99	Inglesa, Espanhola, entre outras.	Windows
Freesr Speech Recognition	Exp.	ND	Inglesa.	Windows
Simon	Sim	0	Inglesa.	Linux, Windows
Web Speech API	Sim	0	Portuguesa e muitas outras.	Todas
Voice Note	Sim	0	Portuguesa e muitas outras.	Todas

Tabela 4.1: Comparação dos *Software Speech-To-Text*

- **Voice Note**⁷ - Esta ferramenta é uma extensão que pode ser instalada no *browser Google Chrome*. Trata-se de uma extensão gratuita e que pode ser usada em Windows, Mac e Linux. No que concerne ao suporte de idiomas, esta extensão, suporta igualmente muitas línguas, sendo que uma delas é a língua Portuguesa.

A Tabela 4.1 apresenta de forma sintetizada os *softwares* estudados e apresentados acima. Ora, de todos eles, os únicos que cumprem os requisitos para este trabalho, ou seja, reconhecer a língua portuguesa e ser uma ferramenta gratuita são a *Web Speech API* e a extensão *Voice Note*. Destas duas ferramentas adequadas ao uso no projeto, surgiu a necessidade de escolher qual a melhor para ser aplicada no nosso projeto.

4.1 *Web Speech API* ou *VoiceNote*?

Para escolher qual destas ferramentas melhor se adequava ao projeto foram realizados vários testes com frases em português para se descobrir qual a ferramenta com melhor performance. As Tabelas 4.2, 4.3 e 4.4 ilustram alguns dos testes realizados com as duas ferramentas. De salientar que cada frase testada foi retirada de um relatório médico, pelo que, a tabela apresenta a frase presente no relatório, bem como, a frase extraída pela *Web Speech API* e pela extensão *VoiceNote* quando é ditada para cada uma delas a frase do relatório.

Como é possível constatar pelas tabelas, os resultados são bastante similares entre as duas ferramentas, ou seja, entre a *Web Speech API* e a extensão *VoiceNote*. De notar que quando comparado com a frase original obtida dos relatórios a performance de cada

⁷<https://voicenote.in>

Relatório	A pele e o tecido celular subcutâneo apresentam aspectos mamográficos normais
WS API	a pele e o tecido celular subcutâneo apresentam aspectos demográficos normais
VoiceNote	a pele e do tecido celular subcutâneo apresento aspectos demográficos normais

Tabela 4.2: Comparação entre *Web Speech API* e *VoiceNote* - Exemplo 1

Relatório	Não se individualizam imagens nodulares que sugiram malignidade, micro-calcificações suspeitas ou outras alterações significativas, em qualquer dos lados
WS API	não consigo visualizar imagens nodulares que sugiro malignidade microcalcificações suspeitas outras alterações significativas em qualquer dos lados
VoiceNote	não consigo visualizar imagens no solares que sugiro malignidade microcalcificações suspeitas outras alterações significativas em qualquer dos lados

Tabela 4.3: Comparação entre *Web Speech API* e *VoiceNote* - Exemplo 2

uma das ferramentas parece apresentar uma performance aceitável. Uma curiosidade que surgiu da comparação destas duas ferramentas foi a tentativa de descobrir se a extensão *VoiceNote* foi implementada recorrendo ao uso da *Web Speech API*. Nós acreditamos que sim, visto nos parecer o mais lógico com base nos resultados, idiomas disponíveis e a natureza da ferramenta, contudo a documentação da extensão não é muito clara em relação a este assunto.

Posto isto, a ferramenta que decidimos usar para incluir no MammoClass foi a *Web Speech API*. A decisão foi simples de tomar pois a *Web Speech API* é uma API, o que por si só permite maior flexibilidade no seu uso. Ou seja, por exemplo, pode ser aplicada em qualquer elemento HTML à escolha, tornando-se assim fácil de incluir numa página Web ao contrário da extensão *VoiceNote* em que não há forma de controlar o que fazer com o texto obtido.

4.2 Problemas inerentes ao uso da *Web Speech API*

Como visto na secção anterior a ferramenta incluída na nova versão do MammoClass foi a *Web Speech API*. Porém a utilização desta API não garante a confidencialidade dos dados, o que é um requisito necessário quando se trata de dados médicos.

Posto isto, surgiu a necessidade de dotarmos o MammoClass de uma ferramenta de *Speech-To-Text* capaz de tratar os dados sem os enviar para servidores externos.

Relatório	No actual estudo, observamos padrão mamográfico de densidades fibroglandulares dispersas, pela pequena quantidade de parênquima mamário
WS API	no atual estudo observamos pedro mamográfico de densidades fibroglandular dispersas pela pequena quantidade de parênquima mamário
VoiceNote	no atual estudo observamos pedro mamográfico de densidades fibroglandular dispersas pela pequena quantidade parênquima mamário

Tabela 4.4: Comparação entre *Web Speech API* e *VoiceNote* - Exemplo 3

Atendendo aos requisitos deste trabalho, ou seja, reconhecer a língua portuguesa e ser uma ferramenta gratuita, como visto anteriormente, não existem ferramentas que resolvam diretamente o problema. Contudo Figueiredo & Dutra (2016) estudaram este problema e apresentaram uma solução que achamos ser capaz de satisfazer a necessidade criada através da *Web Speech API*. Na próxima secção, apresentamos uma extensão do nosso trabalho original Rocha *et al.* (2016) que suporta o Julius⁸ e o Coruja⁹.

4.3 Julius e Coruja

Figueiredo & Dutra (2016) apresentam como ferramenta alternativa ao uso da *Web Speech API* no âmbito de *Speech-To-Text* na área médica, o uso de dois projetos independentes, liderados por instituições académicas estrangeiras e suportados online pela comunidade *open-source*: o *Julius* e o *Coruja*.

O *Julius* é um *software* de *Speech-To-Text*. Originalmente foi desenvolvido por Lee Akinobu do *Nagoya Institute of Technology*, no Japão, funcionando apenas na língua Japonesa. O seu código livre permite que atualmente seja suportado por uma comunidade *online* que contribui para alargar o alcance do programa para outras línguas. Este sistema suporta um vocabulário até cerca de 64.000 palavras, sendo capaz de realizar a descodificação da fala quase em tempo real, mesmo quando está a ser executado num computador pessoal comum. De salientar que o algoritmo do programa assenta em modelos do tipo *Hidden Markov Models*, e incorpora diversas formas de pesquisa, contudo esta ferramenta foi cuidadosamente desenvolvida de forma modular, pelo que é independente de estruturas de modelos, suportando assim diversos tipos de *Hidden Markov Models*. É de ter em consideração que a maior parte da documentação do *Julius* ainda se encontra em língua Japonesa.

O *Coruja* é uma parte integrante e fundamental do projeto FalaBrasil¹⁰. Este *software* foi desenvolvido pelo Laboratório de Processamento de Sinais da Universidade Federal do Pará, no Brasil. O seu desenvolvimento teve como objetivo disponibilizar recursos para a construção de sistemas de *Speech-To-Text* em português do Brasil sendo

⁸http://julius.osdn.jp/en_index.php

⁹<http://www.laps.ufpa.br/falabrasil/reconhecedor.php>

¹⁰<http://www.laps.ufpa.br/falabrasil>

que diversos programas com finalidades diversas já foram elaborados com base neste trabalho. O *Coruja* é composto por uma API desenvolvida em C++, com suporte a *Common Language Runtime* (CLR), para o decodificador do *Julius* (Figueiredo & Dutra, 2016). Contém os modelos acústicos e linguísticos criados com o *Hidden Markov Model Toolkit*¹¹ necessários para o reconhecimento da Língua Portuguesa com pronúncia característica do Brasil. O *Coruja* dispõe ainda de um dicionário fonético (*USPADic*) readaptado de forma a incluir um vocabulário correspondente a 65.532 palavras (máximo aceite pelo *Julius*) com a devida equivalência às respectivas pronúncias em Português do Brasil.

O uso em conjunto destes dois projetos permite obter um sistema de reconhecimento automático de discurso completo em Português do Brasil. Figueiredo & Dutra (2016) criaram um novo dicionário fonético para substituir o que é usado no *Coruja*. O processo de criação deste dicionário passou pela análise de diversos textos médicos elaborados por clínicos na área da oncologia mamária. Os textos foram então processados criando assim um conjunto de termos que garantidamente foram empregados em relatórios oncológicos. Os autores salientam que devido à abundância de profissionais clínicos neste ramo, assim como as diversas regiões de onde são provenientes, bem como outros fatores que tornam a escrita de cada clínico única faz com que a quantidade de palavras usadas em Portugal em relatórios oncológicos seja mais elevada. Tendo então a lista de termos sem repetições os autores procederam para cada termo à criação da sua forma fonética. De salientar, que foram adicionados a este dicionário alguns termos BI-RADS[®] que não apareceram em nenhum dos relatórios mas que são importantes na área da oncologia mamária. Por último, como forma de otimização foram introduzidas certas expressões completas que se evidenciaram pela elevada frequência que apareceram nas palavras testadas pelos autores. Nesta fase o dicionário contém 635 palavras.

Inclusão no MammoClass

Para o projeto elaborado nesta dissertação decidimos que a solução apresentada por Figueiredo & Dutra (2016) seria uma mais valia para o MammoClass.

Posto isto, surgiu a necessidade de adaptar estes dois projetos para funcionamento online. Procedemos à sua instalação no servidor onde o MammoClass está alojado

¹¹<http://htk.eng.cam.ac.uk/>

(o manual de instalação segue em anexo), contudo, a sua instalação no servidor não torna os projetos diretamente integráveis no MammoClass. Para tal, é necessário uma forma de enviar o áudio do utilizador para os dois projetos.

Para resolver esta questão decidimos adaptar o exemplo apresentado por Chris Wilson no seu projeto *Audio Recorder*¹² que faz uso do *plugin Recorder.js*¹³, que se trata de um *plugin* capaz de gravar e exportar o conteúdo produzido pela *Web Audio API*. Esta API foi introduzida juntamente com o HTML5 e trata-se de uma API que permite ao programador processar e sintetizar áudio em aplicações Web. Através desta API podemos fazer com que o *browser* recolha o som do utilizador e o envie para o servidor onde o MammoClass está alojado, som este que irá ser processado pelo Julius e pelo Coruja.

Contudo, o som obtido pelo *browser* e enviado para o servidor não se encontra no formato aceite pelo *Julius*. O *Julius* necessita de um ficheiro *.wav*, com apenas um canal de áudio e com a frequência de amostragem de 22050. Deste modo, após o som estar no servidor e antes de ser processado pelo *Julius* é necessário realizar a conversão desse som para o formato suportado.

Para realizar a conversão desse som recorreremos à ferramenta *SoX*¹⁴. Esta ferramenta é um utilitário multi-plataforma (disponível para Windows, Linux e MacOS) para linha de comandos. Através dela é possível converter vários formatos de ficheiros de áudio em outros formatos. Além de fazer a conversão de ficheiros de áudio este utilitário permite aplicar vários efeitos de som e ainda fazer a reprodução de ficheiros de áudio.

Assim sendo, o fluxo de dados até o som estar preparado para ser enviado para o *Julius* pode ser caracterizado pela Figura 4.1 e consiste em o som ser recolhido pelo microfone através da *Web Audio API*, de seguida o som é enviado para o servidor onde o MammoClass está alojado através de um método POST. De seguida o som é convertido para o formato aceite pelo Julius através da ferramenta *SoX* e finalmente o Julius é invocado com o ficheiro *.wav* resultante.

¹²<https://github.com/cwilso/AudioRecorder>

¹³<https://github.com/mattdiamond/Recorderjs>

¹⁴<http://sox.sourceforge.net/sox.html>



Figura 4.1: Fluxo dos dados para incluir o *Julius* e o *Coruja* no MammoClass

Capítulo 5

Speech-to-Text Interface

Este capítulo explica como foi criada a Interface *Speech-to-Text* para integração no MammoClass.

Como mencionado anteriormente a interface original do MammoClass apenas permite que o utilizador introduza os descritores BI-RADS[®] através de um formulário de inserção manual. Ora, neste trabalho foi construída uma nova interface que permite que o utilizador introduza texto livre numa *text box*, texto este que pode ser escrito ou colado proveniente de um relatório mamográfico.

Esta nova interface permite ainda que o utilizador dite o relatório mamográfico.

Tal como na versão anterior, a nossa implementação foi feita em forma de interface Web. A construção de tal interface foi baseada em *Bootstrap*¹, que se trata de uma *framework* para criar páginas web com um visual moderno e com a principal vantagem da página criada poder ser vista de forma confortável, quer em computadores, quer em dispositivos móveis. Esta confortabilidade advém do facto de através do *Bootstrap* ser possível desenvolver páginas em que os seus componentes escalam e posicionam-se de acordo com a natureza do dispositivo onde a página está a ser visualizada.

¹<http://getbootstrap.com/>

5.1 MammoClass

Esta secção visa apresentar com algum detalhe o que é o MammoClass ² e como este funciona.

O MammoClass permite a classificação de uma mamografia usando um pequeno conjunto de achados mamográficos. Para esta classificação Ferreira (2010); Ferreira *et al.* (2015) usou um *dataset* com 348 casos obtidos entre outubro de 2005 e dezembro de 2007. De forma a encontrar o melhor modelo, foram aplicados sobre estes dados diversos algoritmos de classificação com variação de parâmetros, com o intuito de aprender a partir dos dados. O objetivo destes algoritmos, que foram implementados no WEKA, era a previsão da densidade da massa e a previsão da malignidade. O melhor classificador na tarefa de prever a densidade da massa é baseado em *SVM* (*Support Vector Machine*) e tem uma *accuracy* de 81,3%. No que toca à tarefa de classificação da malignidade o melhor classificador é baseado igualmente em *SVM* e tem uma *accuracy* de 85,6%, com um valor de predição positivo de 85%. Um aspeto importante do MammoClass é a sua capacidade para prever malignidade na ausência do atributo "densidade da massa", uma vez que o modelo consegue prever um valor para este atributo e usá-lo na classificação.

Este modelo foi então disponibilizado online em conjunto com um formulário Web de forma a passar os parâmetros ao modelo. Esta interface Web está representada na Figura 5.1. Como pode ser visto esta interface possui alguns campos que podem ser preenchidos com os valores dos achados mamográficos. O botão "Predict" permite obter um valor de malignidade calculado com base nos valores inseridos.

Queremos salientar que o MammoClass foi um ponto de partida muito importante para o projeto desenvolvido nesta dissertação.

5.2 *Parser* extrator de descritores BI-RADS[®]

A interface elaborada nesta dissertação usa uma adaptação do *parser* elaborado em Cunha (2011); Nassif *et al.* (2012). Esta secção visa apresentar as características desse *parser*.

²<http://cracs.fc.up.pt/nf/mammoclass>

MammoClass

Classification of a mammogram based in a reduced set of mammography findings

To obtain a prediction in terms of malignancy for a certain mass is only necessary to provide the values of the findings, annotated through the Breast Imaging Reporting and Data System (BIRADS), in the form bellow. It is also possible to get a prediction of the attribute *mass density* in case this feature is not known.

The output will indicate the probability of a certain mass being benign or malignant. In the latter case it is suggested that the patient should perform a biopsy. The probabilities are computed using machine learning models built as described in:

- *P.Ferreira, N. A. Fonseca, I. Dutra, R. Woods, and E. Burnside, Predicting Malignancy from Mammography Findings and Surgical Biopsies, submitted.*

Enter Data

Patient's age

Mass size

Breast Composition

Mass shape

Mass clockface location

Mass margins (1)

Mass margins (2)

Mass margins worst

Mass density

Side

Quadrant

Depth

Figura 5.1: Interface do Mammoclass original

O léxico BI-RADS[®] é composto por 43 atributos, contudo os textos médicos possuem alguma variedade de conceitos que não são descritos de maneira uniforme. Segundo Nassif *et al.* (2009) alguns médicos utilizam palavras diferentes para se referirem ao mesmo conceito. Como exemplo temos "igual densidade" e "isodenso", que caracterizam o mesmo atributo BI-RADS[®] relacionado com a densidade, contudo está escrito de forma diferente. De acordo com Cunha (2011) alguns léxicos podem ser descritos de forma ambígua ou até referirem-se a mais do que um conceito. Posto isto na elaboração do *parser* Cunha (2011) teve que conhecer a forma como os textos médicos são escritos e adquirir conhecimento sobre os sinónimos utilizados para descrever alguns atributos BI-RADS[®]. Para isto foi essencial o recurso a uma técnica de radiologia com o intuito de conhecer melhor a linguagem médica relacionada com o cancro da mama e termos BI-RADS[®]. Como estes termos foram criados originalmente em Inglês, Cunha (2011) obteve de uma técnica de radiologia a tradução dos termos em Português.

Para mapear as palavras e as frases do texto em conceitos, não nos podemos guiar apenas pelo léxico. Ora, após possuímos um conhecimento aprofundado sobre os textos, é essencial possuir uma lista de termos bem definida para a extração de dados. Como forma de exemplo temos o conceito "lesão da pele" que é capturado pela presença de ambas as palavras, "lesão" e "pele", desde que estas se encontrem próximas. Posto isto, foi necessário estabelecer a ordem das palavras e em alguns casos o seu grau de proximidade. Em consequência das diferentes formas de como os médicos escrevem a mesma palavra e devido aos erros ortográficos foi definido um conjunto de palavras de forma a ser possível capturar o termo pretendido. Por exemplo, foi definido "adenomegalia", "axila positiva" e "gânglio axilar" para definir o achado associado "Adenopatia Axilar".(Cunha, 2011)

Após a gramática ter detetado a ocorrência de um conceito, é feita a deteção da negação, ou seja, ao percorrer o texto caso a palavra "massa" seja encontrada seguida da palavra "oval", é ativada a regra correspondente à "forma oval". Por outro lado, caso a frase seja constituída por "a massa não é oval" está-se perante um forma negada, logo a regra "forma oval" não é definida.

De acordo com os termos BI-RADS[®] traduzidos para português, Cunha (2011) desenvolveu uma gramática livre de contexto com o intuito de ajudar na extração de atributos BI-RADS[®] de textos clínicos. Inicialmente foi construído um dicionário


```

<Forma> ::= <Farredondada> | <Foval> | <Flobular> | <Firregular>
<Farredondada> ::= <Forma_Expressao>{<ConjCaracteres>} <Farr_Expressao1>
<Farr_Expressao1> ::= arredondado | arredondada | redondo | redonda
<Foval> ::= <Forma_Expressao>{<ConjCaracteres>} <Fov_Expressao1>
<Fov_Expressao1> ::= oval | ovoide | ovalar | alongado | alongada
<Flobular> ::= <Forma_Expressao>{<ConjCaracteres>} <Flob_Expressao1>
<Flob_Expressao1> ::= lobular | polilobular | multilobular
<Firregular> ::= <Forma_Expressao>{<ConjCaracteres>} <Firr_Expressao1>
<Firr_Expressao1> ::= irregular

```

Figura 5.2: Gramática para extração do conceito "Forma"

para cada termo traduzido. Este dicionário pode ser visto na Tabela 5.1 que sumariza a informação apresentada no Capítulo 2.

De seguida foi definida a gramática para a extração dos atributos. Para simplificação alguns símbolos não terminais desta gramática não estão totalmente definidos. Por exemplo, o símbolo não terminal <Caracter> representa qualquer símbolo do teclado. Segundo Cunha (2011), o léxico BI-RADS[®] é composto por termos e construções bem definidos que podem ser facilmente extraídos de relatórios médicos a partir de expressões sem teor sintático ou semântico. Portanto, a gramática definida é baseada em combinações de expressões médicas. Por exemplo, para extrair o conceito "Forma Arredondada", é suficiente encontrar qualquer sinónimo para as palavras "forma" e "arredondada" no texto, possivelmente separadas por um número limitado de palavras. A Figura 5.2 ilustra a gramática usada pelo *parser* para encontrar a característica "forma".

Para a elaboração desta gramática Cunha (2011) utilizou um conjunto de dados contendo registos clínicos e resultados de mamografias de 153 pacientes que foram examinados no período de 2008 a 2009. A extração de termos tem uma taxa de acerto entre os 87% e os 91%.

A inclusão deste *parser* no nosso projeto permite a extração dos termos BI-RADS[®] presentes quer no texto ditado, quer no texto escrito ou copiado, o que irá ser usado no processo de preenchimento automático do formulário da versão do MammoClass elaborada nesta dissertação.

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

Tabela 5.1: Termos BI-RADS[®] e a sua tradução para Português

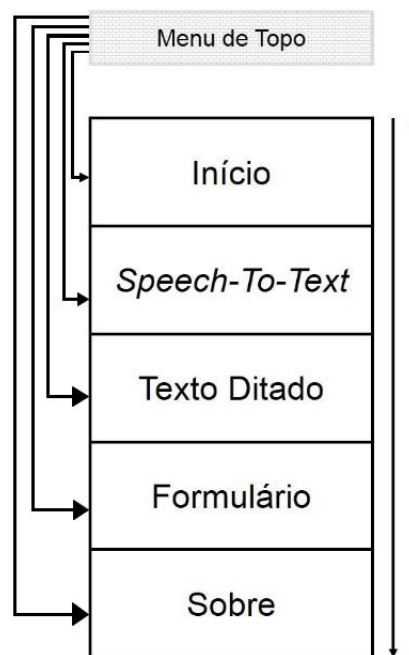


Figura 5.3: Esquema da interface do MammoClass

5.3 Interface Criada

A interface criada é composta por cinco partes distintas que compõem o MammoClass. Estas partes são a interface inicial que é a parte mostrada aquando de uma visita de algum utilizador, a parte de *Speech-To-Text* que permite ao utilizador ditar um relatório médico, a parte de inserção de texto manual de um relatório médico, a parte de preenchimento do formulário com os descritores BI-RADS[®] e por último uma parte com informação sobre o MammoClass. A imagem 5.3 ilustra tal interface. De salientar a presença de um "Menu de topo" que se encontra sempre no topo da interface e que permite a qualquer momento navegar entre cada uma das partes da interface. Posto isto cada uma das seguintes subsecções visa apresentar cada uma das partes da interface que constituem o MammoClass.

5.3.1 Interface Inicial

Quando qualquer utilizador visita a página do MammoClass é lhe apresentado a interface apresentada nas Figuras 5.4 e 5.5, sendo que a Figura 5.4 corresponde à interface



Figura 5.4: Interface apresentada aquando da visita à página do MammoClass - Computador e *Tablet*

visualizada por utilizadores oriundos de um computador ou *tablet* enquanto a Figura 5.5 representa a interface visualizada por utilizadores oriundos de um *Smartphone*

Ora na parte superior da Figura 5.4, como referido anteriormente, podemos ver um menu com diversos textos de navegação. Este "Menu de topo" é representado na figura 5.6. Este "Menu de topo" é mantido no topo da interface aquando da navegação para outras partes da página. Ao manter este menu no topo o utilizador pode sempre que queira, de formar rápida, navegar para qualquer outra parte da interface. No que concerne ao utilizador proveniente de um *Smartphone*, este tem à sua disposição de igual forma um "Menu de topo" para tal basta carregar no botão representado por três barras horizontais, este botão pode ser visto no canto superior direito da Figura 5.5. O aspeto deste menu de topo pode ser visto na Figura 5.7.

Para cada uma das plataformas de visualização ao clicar em "MAMMOCLASS V2" o utilizador é conduzido de novo para a interface inicial da página, ou seja, a interface representada pelas Figuras 5.4 e 5.5, respetivamente. Por sua vez ao clicar no texto "STT" o utilizador é conduzido para a parte da interface que lida com a entrada de texto por voz, ou seja, texto que é ditado pelo utilizador. Ao clicar em "DATA MANUALLY" o utilizador é conduzido para a parte da interface responsável pela introdução de texto manual. Por sua vez, ao clicar em "FILL FORM" o utilizador

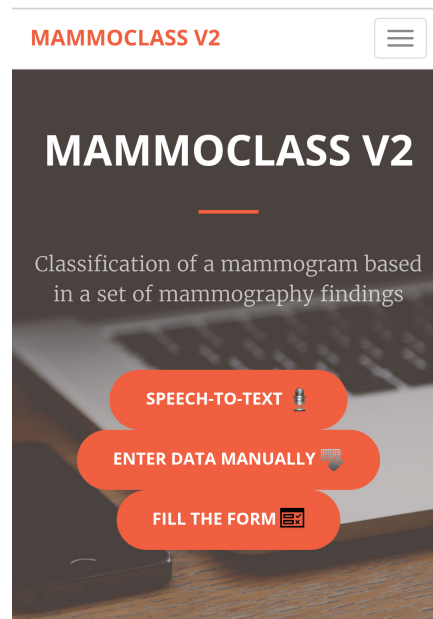


Figura 5.5: Interface apresentada quando da visita à página do MammoClass - *Smartphone*



Figura 5.6: Menu presente sempre no topo da interface - Computador e *Tablet*

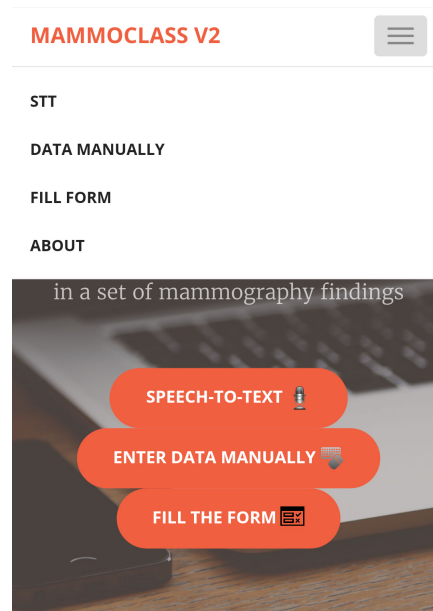


Figura 5.7: Menu acessível sempre no topo da interface - *Smartphone*

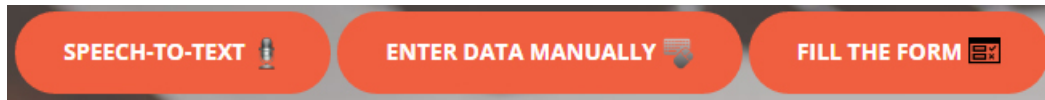


Figura 5.8: Botões de navegação fácil presentes na interface inicial

é redirecionado para a parte da interface que contem o formulário de descritores BI-RADS[®] onde o utilizador pode preencher os descritores e obter o valor de predição da malignidade associada aos descritores. Por último ao clicar em "ABOUT" o utilizador é conduzido para uma parte da interface com informação sobre o MammoClass.

No centro desta interface inicial está como pode ser visto nas Figura 5.4 e 5.5 o nome do projeto, ou seja, "MammoClass" e uma pequena descrição da sua principal função: *Classification of a mammogram based in a set of mammography findings*. Para além deste conteúdo existem ainda botões de navegação que permitem ao utilizador navegar na página para a parte pretendida. No que toca ao utilizador do computador ou *tablet* estes botões podem ser vistos na figura 5.8, já no que respeita ao utilizador de *smartphone* os botões são os mesmos e apenas estão dispostos de forma diferente como pode ser visto na Figura 5.5. Assim sendo, o botão com o texto "SPEECH-TO-TEXT" permite ao utilizador navegar para a parte da interface que lida com a entrada de texto por voz. O botão com o texto "ENTER DATA MANUALLY" permite, como o nome indica, que o utilizador seja conduzido para a parte da interface responsável pela introdução de texto manual. Por sua vez, ao clicar no botão com o texto "FILL THE FORM" o utilizador é redirecionado para a parte da interface que contem o formulário de descritores BI-RADS[®] para previsão de malignidade. Como seria de esperar visto tratar-se de uma página Web, qualquer parte desta interface pode ser acedida fazendo *scroll* com o rato ou usando a barra de navegação do *browser* caso estejamos num computador, ou fazendo o *scroll* com o dedo caso o utilizador esteja a utilizar um *tablet* ou um *smartphone*.

5.3.2 Interface para *Speech-To-Text*

A parte da interface responsável pela componente *Speech-To-Text* está representada na Figura 5.9. Quando visualizada por um utilizador proveniente de um *Smartphone* esta parte da interface tem o aspeto apresentado na Figura 5.10. Esta parte da interface contem uma imagem de um microfone para fazer alusão à sua função, ou



Figura 5.9: Interface responsável pela componente *Speech-To-Text* - Computador e *Tablet*



Figura 5.10: Interface responsável pela componente *Speech-To-Text* - *Smartphone*

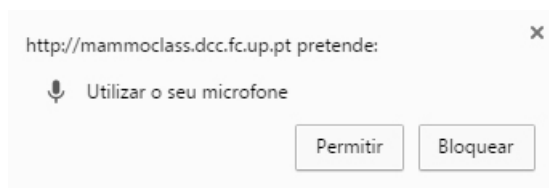


Figura 5.11: *Pop-up* do *Google Chrome* para permissão de recolha de voz

seja, capturar texto ditado pelo utilizador. Contem também as duas ferramentas de *Speech-To-Text* que o MammoClass suporta, permitindo assim ao utilizador escolher qual pretende utilizar através de *radio buttons*. Nesta parte da interface estão ainda dois botões, sendo estes o botão "START VOICE RECOGNITION" e o botão "END". Ao clicar no botão "START VOICE RECOGNITION" é iniciado o reconhecimento de discurso com a ferramenta de *Speech-To-Text* selecionada. Neste ponto queremos chamar a atenção de um aspeto importante, ora, por razões de segurança, antes de ser iniciada qualquer recolha de voz por parte de cada uma das ferramentas, o utilizador deve permitir tal recolha. Por isso, ao clicar em "START VOICE RECOGNITION" surge no *browser* a pergunta ilustrada na figura 5.11 na qual o utilizador deve carregar em "Permitir" para que a captura de texto seja bem sucedida.

Ora, estando a gravação no estado de iniciada a imagem do microfone muda de forma a permitir ao utilizador saber que a sua voz está a ser recolhida. O aspeto desta parte da interface quando a recolha de voz é feita através de computador ou *tablet* encontra-se ilustrado pela Figura 5.12. Na visualização através de um *Smartphone* o aspeto da interface está ilustrado na Figura 5.13.

Quando o utilizador terminar de realizar o ditado do texto e pretender finalizar a recolha de voz para o seu texto ser processado, deve carregar no botão "END". Ao carregar neste botão o utilizador para além de terminar a recolha de voz é redirecionado para a parte da interface com o formulário de descritores BI-RADS[®] já preenchidos com os descritores extraídos do texto ditado.

5.3.3 Interface para digitar texto

Esta parte da interface é responsável por lidar com o digitar de texto manual ou a inserção de texto copiado de um relatório. Ora, no caso do utilizador ser proveniente de um computador ou *tablet* esta componente é representada pela Figura 5.14. No

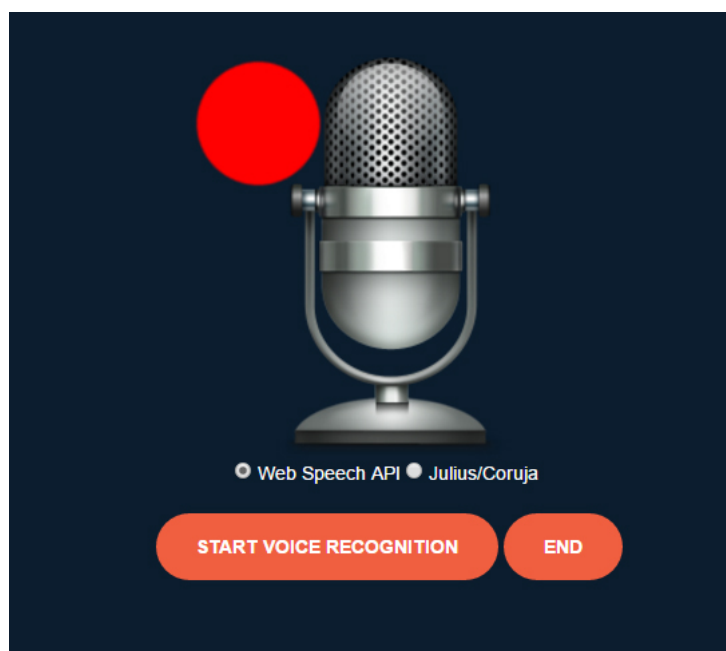


Figura 5.12: Interface responsável pela componente *Speech-To-Text* quando gravação - Computador e *Tablet*



Figura 5.13: Interface responsável pela componente *Speech-To-Text* quando gravação - *Smartphone*

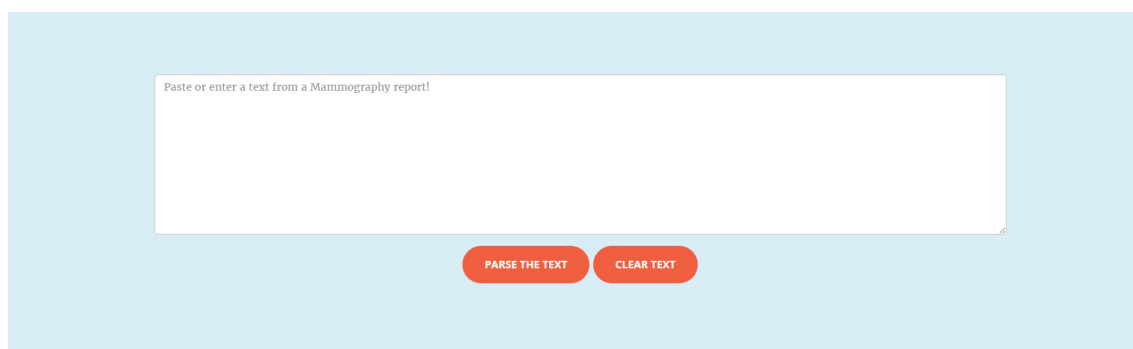


Figura 5.14: Interface responsável pela componente de texto ditado - Computador e *Tablet*

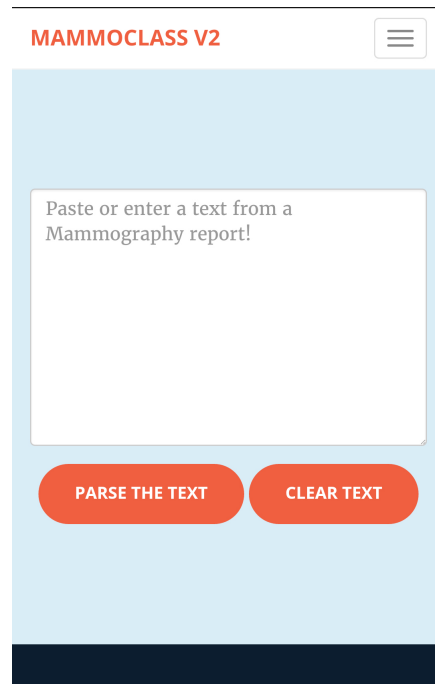
caso da proveniência do utilizador recair no *Smartphone* o aspeto desta componente pode ser visto na Figura 5.15

Como podemos ver esta componente contem um campo onde pode ser escrito ou colado o texto proveniente de um relatório mamográfico. Juntamente com este campo existem dois botões, o botão "CLEAR TEXT" que, como o nome indica, serve para limpar o texto presente no campo de inserção de texto e o botão "PARSE THE TEXT" que, por sua vez, faz com que a interface processe o texto digitado ou copiado e de seguida, redireciona o utilizador para a parte da interface com o formulário de descritores BI-RADS[®] já preenchidos com os descritores extraídos do texto, que foi escrito ou colado no campo de inserção de texto.

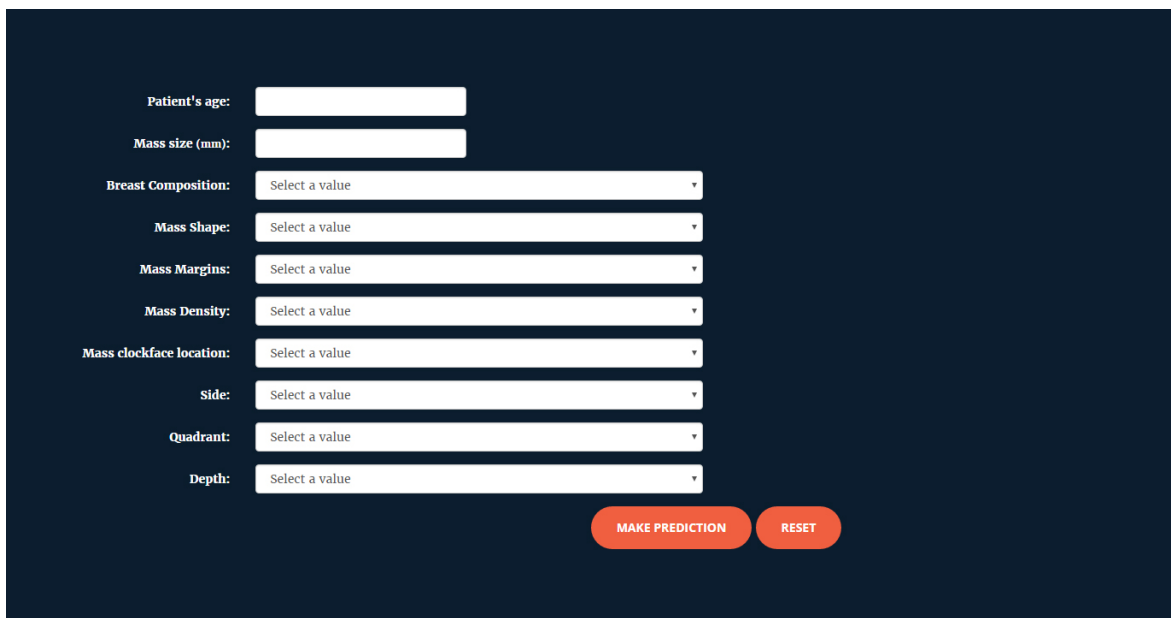
5.3.4 Interface com o formulários de descritores BI-RADS[®]

Nesta componente da interface está presente o formulário com os descritores BI-RADS[®]. Este formulário caso o utilizador esteja a usar um computador ou *tablet* tem o aspeto que pode ser visto na Figura 5.16. Caso o utilizador esteja a usar um *Smartphone* o aspeto da interface pode ser visto na Figura 5.17. Devido aos diversos campos presentes no formulário o utilizador de *Smartphone* necessita de fazer *scroll* no ecrã de forma a percorrer todo o formulário. Ora a Figura 5.17a ilustra o aspeto do início do formulário enquanto que a figura 5.17b ilustra o fim desta parte da interface.

Ora esta componente da interface foi reformulada com base na interface do Mammo-Class existente. Comparativamente com a versão anterior, para além das alterações a nível visual foi alterada a ordem dos descritores no formulário. Foi ainda eliminado



The screenshot shows the MAMMOCLASS V2 application on a smartphone. At the top, the title "MAMMOCLASS V2" is displayed in red, next to a hamburger menu icon. Below the title is a large white text input area with the placeholder text "Paste or enter a text from a Mammography report!". At the bottom of the input area, there are two red buttons: "PARSE THE TEXT" and "CLEAR TEXT".

Figura 5.15: Interface responsável pela componente de texto ditado - *Smartphone*

The screenshot shows a form for BI-RADS descriptors on a dark blue background. The form includes the following fields:

- Patient's age:
- Mass size (mm):
- Breast Composition:
- Mass Shape:
- Mass Margins:
- Mass Density:
- Mass clockface location:
- Side:
- Quadrant:
- Depth:

At the bottom right, there are two red buttons: "MAKE PREDICTION" and "RESET".

Figura 5.16: Interface com o formulários de descritores BI-RADS[®] - Computador e *Tablet*

The figure consists of two side-by-side screenshots of a mobile application interface titled "MAMMOCLASS V2".

(a) Início formulário: This screenshot shows the beginning of the form. It features a dark blue background with white text and input fields. The fields are: "Patient's age:" (text input), "Mass size (mm):" (text input), "Breast Composition:" (dropdown menu with "Select a value"), "Mass Shape:" (dropdown menu with "Select a value"), "Mass Margins:" (dropdown menu with "Select a value"), and "Mass Density:" (dropdown menu with "Select a value").

(b) Final do formulário: This screenshot shows the end of the form. It features the same dark blue background. The fields are: "Mass Entrance location:" (dropdown menu with "Select a value"), "Side:" (dropdown menu with "Select a value"), "Quadrant:" (dropdown menu with "Select a value"), and "Depth:" (dropdown menu with "Select a value"). At the bottom, there are two red buttons: "MAKE PREDICTION" and "RESET".

Figura 5.17: Interface com o formulário de descritores BI-RADS[®] - *Smartphone*

um dos campos *Mass Margins* pois a sua utilização era redundante em termos de classificação.

O botão "RESET" limpa todos os valores dos descritores. Este botão é útil pois caso o utilizador tenha chegado à conclusão que deve mudar todos os descritores que inseriu pode assim através deste botão realizar de forma fácil um *reset* ao formulário. O botão "MAKE PREDICTION" comunica ao sistema que deve ser feita uma previsão de malignidade com base nos descritores BI-RADS[®] presentes no formulário. Caso o preenchimento de algum descritor esteja em falta, foi criado um aviso de alerta emitido pelo *browser* indicando que todos os campos do formulário devem estar preenchidos antes que uma predição possa ser feita. Este aviso de alerta pode ser visualizado na Figura 5.18.

Por sua vez, quando é realizada uma predição a parte da interface que agrega o questionário fica com o formato que se pode observar na Figura 5.19.

Queremos também salientar que quando o utilizador usa uma das outras formas de inserir os descritores, ou seja, ditando um relatório ou digitando o relatório, foi implementada uma interface de avisos para o utilizador no caso do *parser* não extrair

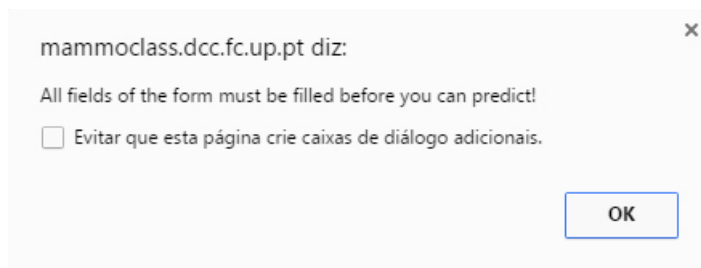
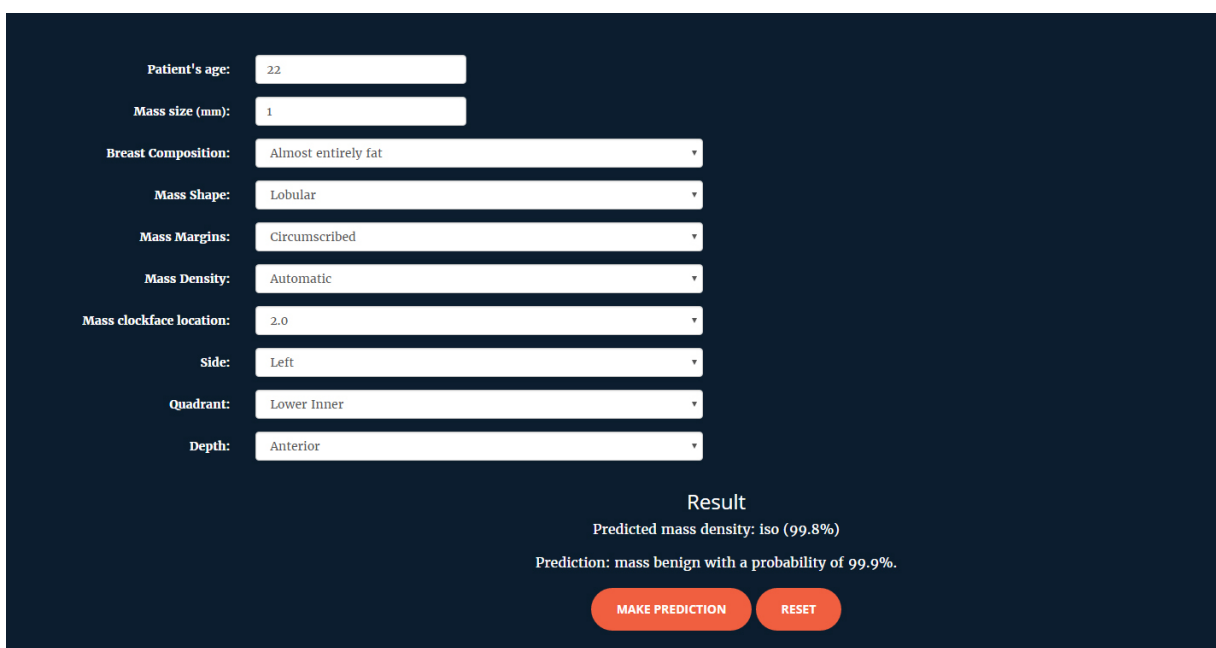


Figura 5.18: Aviso emitido quando formulário está incompleto.



Patient's age:	22
Mass size (mm):	1
Breast Composition:	Almost entirely fat
Mass Shape:	Lobular
Mass Margins:	Circumscribed
Mass Density:	Automatic
Mass clockface location:	2.0
Side:	Left
Quadrant:	Lower Inner
Depth:	Anterior

Result
Predicted mass density: iso (99.8%)
Prediction: mass benign with a probability of 99.9%

MAKE PREDICTION RESET

Figura 5.19: Interface do formulário após uma predição

Figura 5.20: Interface do formulário com avisos despoletados pelo *parser* - Computador e *Tablet*

determinado atributo, ou de o *parser* não ter encontrado no texto enviado tal atributo. Esta interface de avisos foi considerada útil pois alerta o utilizador para a falta de descritores preenchidos. Esta interface pode ser vista na Figura 5.20 e na Figura 5.21, sendo que a Figura 5.20 representa interface mostrada ao utilizador proveniente de computador e *tablet* enquanto que a Figura 5.21 representa a interface mostrada ao utilizador proveniente de *Smartphone*. Como é possível de ver, a cor vermelha encontram-se os avisos relativos à não extração de tal descritor BI-RADS[®] pelo *parser*, enquanto que, a cor verde podem ser encontrados os avisos inerentes à falta do respetivo descritor BI-RADS[®] no texto digitado ou ditado.

5.3.5 Interface de informação relativa ao projeto

Esta parte da interface é a última parte da página. Nesta parte, caso o utilizador seja proveniente de um computador ou de um *tablet* o seu aspeto pode ser vista na Figura 5.22, caso o utilizador seja proveniente de um *Smartphone* o aspeto da interface pode ser visto na Figura 5.23 onde mais uma vez devido á extensão desta parte da interface, estes utilizadores necessitam de fazer *scroll* na página para percorrerem toda a parte da interface. O conteúdo é mostrado assim nas Figuras 5.23a e 5.23b. Ora, neste parte da interface é apresentado um *disclaimer*, bem como a referência para o artigo

MAMMOCLASS V2

Patient's age:

The parser does not extract this finding

Mass size (mm):

This finding does not exist in the text

Breast Composition:
Select a value
The parser does not extract this finding

Mass Shape:
Select a value
This finding does not exist in the text

Mass Margins:
Select a value
This finding does not exist in the text

Figura 5.21: Interface do formulário com avisos despoletados pelo *parser* - *Smartphone*

original do Mammoclass. Por último são apresentados os contactos dos principais intervenientes no projeto.

5.4 Explicação e Esquematização da nova Ferramenta

Como visto nas secções anteriores, para além das mudanças gráficas feitas em relação ao MammoClass original, ao longo desta dissertação, o MammoClass evoluiu para um sistema que permite a introdução de dados através do discurso. A implementação de tal característica segue o gráfico de fluxo representado na Figura 5.24

De salientar que desenvolvemos para o novo MammoClass um *script* em *JavaScript* que trata de toda a parte responsiva da página, isto é, que trata de todos os *handlers* dos botões e que quando necessário comunica com o servidor onde o MammoClass está alojado. Cada um dos pontos seguintes representa uma etapa do fluxo de informação para o funcionamento da componente *Speech-to-text*.

- **Som traduzido em texto pela API** - Esta é a primeira etapa da componente *Speech-to-text*. Como explicado anteriormente, quando o utilizador pressiona o



Figura 5.22: Interface relacionada com a informações sobre o projeto - Computador e *Smartphone*

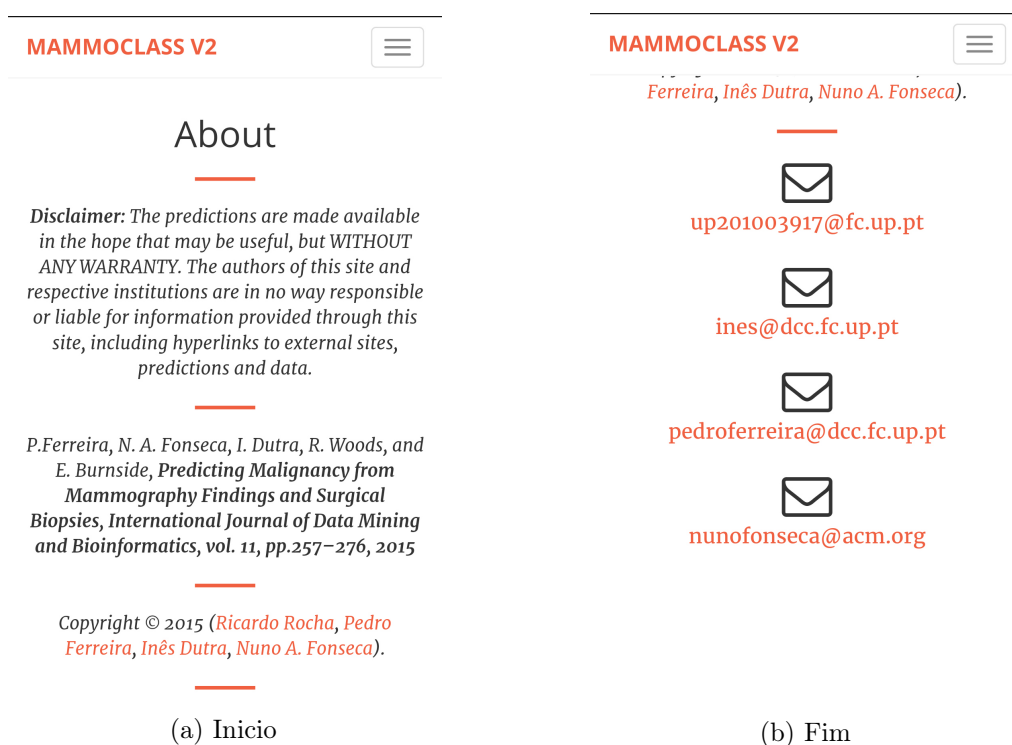


Figura 5.23: Interface relacionada com a informações sobre o projeto - *Smartphone*

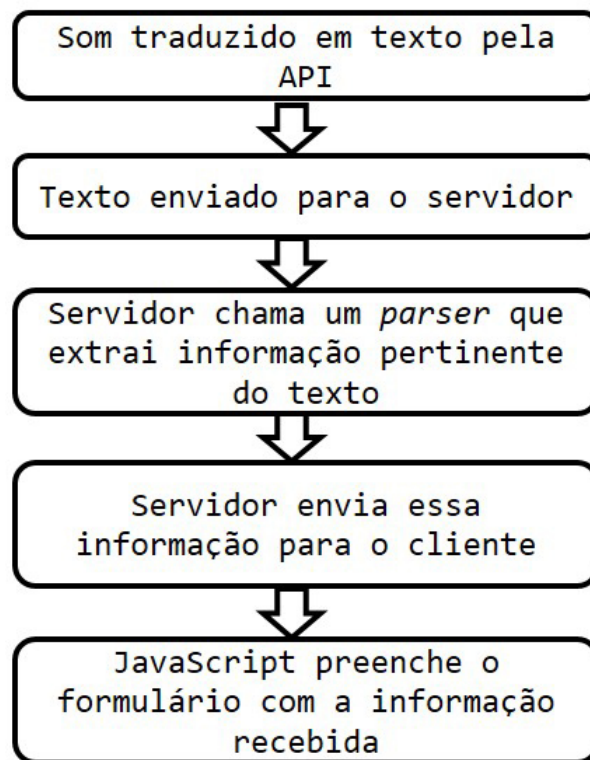


Figura 5.24: Gráfico de Fluxo do funcionamento da componente *Speech-to-text*

botão "START VOICE RECOGNITION" visível nas Figuras 5.9 e 5.10 o *script* em *Javascript* verifica qual a ferramenta de *Speech-To-Text* está selecionada. Caso seja a *Web Speech API*, invoca o método "recognition.start()" pertencente a esta API. Este método ordena a API que comece a transcrever o texto falado para texto escrito. De salientar que tal como mostrado na Figura 5.11 para que a API tenha acesso ao som do microfone o utilizador deve carregar no botão para permitir tal acesso. A partir desse momento a API encontra-se a recolher todo o som que o utilizador dite. Caso a ferramenta seja o Julius/Coruja o som é captado como descrito anteriormente. Por sua vez, quando o utilizador pressiona o botão "END", caso a ferramenta selecionada seja a *Web Speech API* o nosso *script Javascript* invoca o método "recognition.end()" dessa API. Este método ordena a API que devolva ao nosso *script* a sua interpretação do texto ditado pelo utilizador no formato texto. No caso da ferramenta escolhida ser o Julius/Coruja a captura de som é parada e é recebido de igual forma a representação textual do ditado.

- **Texto enviado para o servidor** - Nesta segunda etapa da componente *Speech-To-Text*, tal como o nome indica, o texto é enviado para o servidor em texto livre. Recordando da etapa anterior, neste momento temos o texto ditado pelo utilizador em formato texto no nosso *script* associado ao MammoClass. Assim sendo, é então criado um objeto do tipo "HTTPRequest" que inicia a comunicação com o servidor onde está alojado o MammoClass e que possui um *script CGI* pronto a ser executado. Para esse *script* é então enviada, através de um método POST, a representação em formato texto do discurso ditado pelo utilizador. O nosso *script JavaScript* fica assim à espera de uma resposta do servidor que será recebida numa das etapas seguintes.
- **Servidor chama um *parser* que extrai informação pertinente do texto** - Esta é a terceira etapa da componente *Speech-To-Text*. Nesta etapa o *script cgi* que recebeu o texto vindo do cliente, passa esse texto a um *parser* adaptado do *parser* implementado em Cunha (2011); Nassif *et al.* (2012). Este *parser* extrai os atributos BI-RADS[®] presentes no texto recebido, gerando assim uma tabela booleana com os termos que encontrou.
- **Servidor envia essa informação para o cliente** - Esta trata-se da quarta etapa da componente *Speech-To-Text* e é responsável pelo envio da resposta ao método POST que aguarda uma resposta desde a segunda etapa. Esta resposta trata-se da tabela booleana extraída pelo *parser* enviada no formato de texto livre.
- **JavaScript preenche o formulário com a informação recebida** - Esta é a quinta e última etapa da componente *Speech-to-text*. Nesta componente o *script JavaScript* recebe como resposta ao método POST invocado na segunda etapa a tabela booleana extraída na terceira etapa. Para tratar a informação recebida foi implementado um algoritmo que realiza o *parsing* da tabela de forma a preencher o formulário da página de acordo com os descritores BI-RADS[®] extraídos. Sendo assim, para cada campo do formulário foi estabelecida uma correspondência com o índice da tabela com o intuito de saber a que opção de cada campo do formulário corresponde cada valor da tabela.

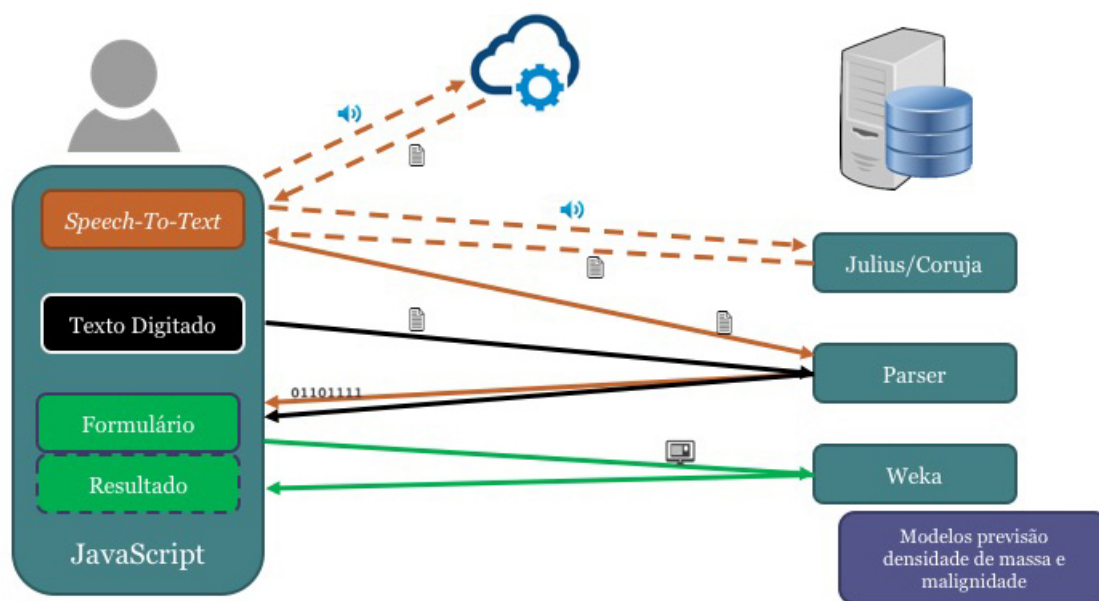


Figura 5.25: Esquemática de toda a ferramenta

Para além de dados introduzidos através do ditar do texto, foi, como visto anteriormente, criada a possibilidade do utilizador digitar ou colar texto proveniente de um relatório nesse caso o texto entra na segunda etapa do fluxo apresentado anteriormente.

Quando o formulário se encontra totalmente preenchido e o utilizador carrega no botão para realizar a previsão de malignidade, o formulário é enviado para um `script` PHP no servidor através de um método POST. Do lado do servidor, este `script` chama o WEKA com os modelos de classificação de forma a obter os resultados correspondentes aos campos do formulário.

Todas estas implementações são exemplificadas pela Figura 5.25. Podemos observar do lado esquerdo todas as componentes mostradas ao utilizador e do lado direito o servidor onde o MammoClass está alojado. No servidor está assim presente a nossa instalação do *Julius* e do *Coruja*, a nossa adaptação do *parser*, bem como a instalação do *Weka* com os modelos de previsão.

Estes são os pontos principais de todo o funcionamento da interface *Speech-To-Text*.

Capítulo 6

Metodologia Experimental

Neste capítulo descrevemos as experiências realizadas, métricas de avaliação e a metodologia experimental utilizada para a avaliação da nova plataforma MammoClass.

Como visto no capítulo anterior a Tabela 5.1 apresenta alguns exemplos dos termos BI-RADS[®], bem como a sua tradução para a língua portuguesa. O léxico BI-RADS[®] (em português) é composto por 86 termos. Assim sendo, para avaliarmos a performance da *Web Speech API* para detetar esses 86 termos BI-RADS[®] realizamos testes com algumas pessoas para ver qual a performance demonstrada. Também fizemos um teste reduzido com o Julius/coruja. Na secção seguinte serão apresentados os materiais necessários para tal avaliação.

6.1 Materiais e Métodos

Para realizar as experimentações à *Web Speech API* recorreremos a quatro indivíduos, dois do sexo masculino e dois do sexo feminino. Estes indivíduos são nativos de Portugal e a língua falada por cada um deles é a Língua Portuguesa de Portugal. Em termos de nível de formação este é semelhante contudo não são especialistas na área da radiologia. Cada um destes indivíduos ditou os 86 termos BI-RADS[®] para uma interface de teste que implementamos ¹. Esta interface usa a *Web Speech API* da mesma forma que esta é usada no MammoClass, só que ao contrário de enviar o texto

¹<http://www.alunos.dcc.fc.up.pt/~up201003917/SpeechToText.html>



Figura 6.1: Interface usada para as Experiências



Figura 6.2: Interface usada para as Experiências após o ditado de um termo

para o servidor, mostra logo ao utilizador qual o resultado do ditado extraído pela API. O aspeto desta interface de teste pode ser visto na Figura 6.1. Por sua vez, o aspeto desta interface aquando a realização de uma experiência pode ser visto na Figura 6.2, onde foi ditado para a API o termo "forma redonda". É possível observar que a API converteu o som para o texto "forma redonda", sendo também possível de ver a confiança com que a *Web Speech API* extraiu o termo. O valor de confiança varia entre 0 e 1, sendo que a API retorna o valor 1 quando está 100% confiante que determinado termo é correto. De salientar que a área de mensagem serve para *debug* uma vez que apresenta alertas de quando o reconhecimento inicia e termina.

Para além dos testes aos 86 termos BI-RADS[®], foram ainda, realizadas experiências com 67 relatórios médicos completos referentes a mamografias realizadas no Centro Hospitalar São João (CHSJ), Porto, Portugal, entre 2008 e 2009. Este trabalho foi

aprovado pelo comitê de ética do CHSJ (referência CES 42-15). Esta experiência foi feita apenas com a *Web Speech API*.

6.2 Experiências

Em termos de testes realizados à *Web Speech API* podemos dividi-los em duas experiências separadas. A Experiência 1 consiste na realização de um ditado dos 86 termos BI-RADS[®] por quatro indivíduos, dois do sexo masculino e dois do sexo feminino. A Experiência 2, por sua vez, consiste nos testes realizados com os relatórios médicos de mamografias e foi realizada apenas por um indivíduo do sexo feminino.

6.2.1 Experiência 1 - Teste aos termos BI-RADS[®]

Nesta experiência cada um dos 86 termos BI-RADS[®] foi ditado individualmente para a interface de teste mostrada na Figura 6.1. Cada termo foi ditado de forma pausada. Os termos foram ditados cada um por quatro indivíduos, dois do sexo masculino e dois do sexo feminino. Cada um deles ditou os termos para três dispositivos de captura de som diferentes:

- Um portátil ao qual foi conectado um microfone externo da marca *NGS*
- O mesmo portátil usando o microfone embutido
- Um *Smartphone*

Ao longo desta dissertação etiquetamos as duas pessoas do sexo masculino como as pessoas A e B. Por sua vez os indivíduos do sexo feminino foram etiquetados como as pessoas C e D.

Para esta experiência classificamos cada resultado retornado pela API como:

- **Correto (C)** - Se o termo original e o termo retornado pela *Web Speech API* são exatamente iguais.

- **Quase Correto (QC)** - Se o termo original e o termo retornado pela *Web Speech API* são "quase" iguais. Mais especificamente, o termo retornado pela API é considerado quase correto se as palavras apenas variam em questões de género, número, grau ou diferente ortografia. Como exemplo temos o termo original "distorção arquitetural" em que a API retornou "distorção arquetural". Como pode ser visto a única diferença entre estas duas palavras é a falta do carácter *c* que foi removido em algumas palavras do novo acordo ortográfico estabelecido entre países que falam a língua portuguesa.
- **Incorreto (I)** - Se o termo original e o termo retornado pela *Web Speech API* são completamente diferentes.

Queremos salientar desde já que consideramos os termos *Quase corretos* como resultados aceitáveis uma vez que podem ser automaticamente corrigidos. Já temos uma implementação preliminar desta correção. Por sua vez os termos *Incorretos* não podem ser corrigidos e são vistos como fonte de erros.

De forma resumida o objetivo desta experiência visa compreender a performance da *Web Speech API* no que concerne à interpretação de termos BI-RADS[®]. Em anexo a esta dissertação está uma tabela com todos os termos testados pelos quatro indivíduos em experiência com a *Web Speech API*.

Relativamente ao Julius/Coruja, só fizemos a Experiência 1, e com um único indivíduo por limitações de tempo.

6.2.2 Experiência 2 - Teste aos relatórios médicos

O objetivo desta experiência é comparar o *output* do *parser* quando o relatório é ditado com o *output* do *parser* quando o texto é copiado e colado de um relatório médico. Para esta experiência o ditado foi realizado apenas por uma pessoa.

Como visto anteriormente o *parser* não é capaz de extrair todos os descritores BI-RADS[®] que compõem o formulário do MammoClass, sendo assim, podemos resumir o objetivo desta experiência em testar se existe alguma perda de performance do MammoClass, quando o utilizador opta por ditar o texto em vez de o escrever ou copiar de um relatório médico.

Capítulo 7

Resultados e Discussão

Neste capítulo serão analisados os resultados obtidos no que concerne à componente *Speech to text* da nova versão do MammoClass desenvolvida nesta dissertação. Os resultados serão apresentados por experiências.

Tal como visto no capítulo anterior, para avaliar a performance da componente *Speech to text* a Experiência 1 consistiu no ditar dos 86 termos BI-RADS[®] de forma individual. Por sua vez, a Experiência 2, consistiu no ditar de relatórios médicos com o intuito de comparar o resultado da extração realizada pelo *parser* quando o texto é proveniente de ditado ou de inserção manual.

De seguida são apresentados os resultados para a Experiência 1, ou seja, os resultados de ditar os 86 termos BI-RADS[®] de forma individual. Para esta experiência reportamos resultados quer da *Web Speech API* quer do Julius/coruja.

7.1 Experiência 1 - Teste aos termos BI-RADS[®]

Como visto no capítulo anterior, após o ditado de um termo, o resultado retornado pela *Web Speech API* pode ser classificado como correto, quase correto ou incorreto. Para debatermos os resultados obtidos, decidimos avaliar a performance da *Web Speech API* de acordo com duas dimensões, a performance por pessoa, bem como a performance por tipo de equipamento.

Assim sendo, os resultados apresentados na Tabela 7.1 mostram as percentagens de termos corretos, quase corretos e incorretos para cada experiência individual de cada pessoa. Esta percentagem é calculada através da fórmula

$$\frac{\text{número de termos da categoria a considerar}}{\text{número total de termos testados}}$$

Como exemplo, podemos observar que a Pessoa A quando usou o portátil com um microfone externo obteve uma taxa de termos corretos de 67,4%, uma taxa de termos quase corretos de 8,2% e uma taxa de termos incorretos de 24,4%. Juntando as taxas de termos corretos e quase corretos (e aqui assumindo que corrigimos automaticamente os termos quase corretos), a Pessoa A obtém uma taxa de acertos de 75,6%. Esta pessoa quando usou o portátil com o microfone incorporado, obteve uma taxa de termos corretos de 68,6%, uma taxa de termos quase corretos de 9,3%, bem como uma taxa de termos incorretos de 22,1%. Se juntarmos as taxas de termos corretos com a taxa de termos quase corretos, obtemos assim, uma taxa de acertos de 77,9%. Por último, esta mesma pessoa quando usou o *Smartphone*, obteve uma taxa de termos corretos de 69,8%, uma taxa de termos quase corretos de 8,1%, bem como uma taxa de termos incorretos de 22,1%. Ora, juntando as taxas de termos corretos e quase corretos a pessoa A obteve uma taxa de acerto de 77,9%.

Para cada um dos dispositivos usados, calculamos a média de cada uma das pessoas (linha etiquetada com Média na Tabela 7.1). Como é possível de ver, se consideramos a taxa de termos corretos o dispositivo com maior taxa é o portátil com microfone externo, dispositivo este que apresenta uma taxa de acerto de 70,1%. Segue-se o portátil com microfone interno que apresenta uma taxa de acerto de 69,2%. Por último surge o *Smartphone* com uma taxa de acerto de 68,9%. Contudo, esta classificação varia caso consideremos a junção da classe de termos corretos com a classe de termos quase corretos. Ora em tal situação, o dispositivo com melhor performance continua a ser o portátil com microfone externo, apresentando uma taxa de acerto de 77,6%, seguido do *Smartphone* com uma taxa de acerto de 75,9%. Em último lugar surge o portátil com microfone interno que apresenta uma taxa de acerto de 74,4%.

Através da Tabela 7.2 é possível observar a performance média por pessoa. A pessoa A em média, acertou em 68,6% dos termos, teve em média 8,5% dos termos classificados como quase corretos e uma média de 22,9% termos classificados de forma incorreta. Se considerarmos a junção das classes de termos corretos e quase corretos temos que

Dispositivo	Pessoa	C(%)	QC(%)	I(%)	C+QC(%)
Portátil com mic externo	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
	Média	70,1	7,5	22,4	77,6
Portátil com mic interno	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
	Média	69,2	5,5	25,3	74,7
<i>Smartphone</i>	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
	Média	68,9	7,0	24,1	75,9

Tabela 7.1: Performance por pessoas

Pessoa	C(%)	QC(%)	I(%)	C+QC(%)
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

Tabela 7.2: Médias por pessoas

a pessoa A em média acertou em 77,1% dos termos. Como é possível constatar pela Tabela 7.2, em relação aos termos corretos, a pessoa com melhor média de acertos é a pessoa B (75,6%), seguindo-se da pessoa C (69,0%), da pessoa A (68,6%) e por último da pessoa D (64,3%). Já se consideramos a junção da classe de termos corretos e quase corretos a pessoa com melhor média de acertos é mais uma vez a pessoa B (81,4%), seguindo-se da pessoa A (77,1%), da pessoa C (76,0%) e por último da pessoa D (69,8%).

Valores neste intervalo foram reportados em outros trabalhos em que os relatórios médicos foram ditados.

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 Distorsão 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.*

Figura 7.1: Exemplos de relatórios médicos

No que diz respeito ao Julius/Coruja, a taxa de acerto utilizando o dicionário original foi de aproximadamente 46,5%. Isto acontece porque o dicionário foi definido com a fonética brasileira Figueiredo & Dutra (2016). Após a redefinição do dicionário, a taxa de acerto subiu para 79,1% de termos corretos (C). A taxa de termos quase corretos (QC) subiu para 81,4%, que é uma taxa comparável com a da *Web Speech API*.

7.2 Experiência 2 - Teste aos relatórios médicos

Neste experiência comparamos o *output* do *parser* quando o relatório é ditado com o *output* do *parser* quando o texto é copiado e colado de um relatório médico. Como visto no capítulo anterior, o ditado nesta experiência foi realizado apenas por uma pessoa.

A Figura 7.1 mostra exemplos de relatórios médicos escritos na língua portuguesa. O primeiro relatório fala sobre uma massa de 2 centímetros no quadrante superior esquerdo da mama, com achados suspeitos de malignidade. A categoria final BI-RADS[®] atribuída é 5. Por sua vez, o segundo relatório descreve uma mama com densidade gorda predominante, com distribuição normal de elementos fibroglandulares, trauma bem conhecido na direita, distorção do estoma na mama direita com 30 milímetros de tamanho.

Para realizarmos esta experiência cada relatório foi colado na interface de introdução de texto manual e depois o mesmo relatório foi ditado.

Para todos os 67 relatórios que testamos o *parser* foi capaz de extrair os mesmos descritores BI-RADS[®] relevantes, apenas com a exceção do tamanho das massas.

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*

Figura 7.2: Relatórios reconhecidos pela *Web Speech API*

Isto deve-se ao facto dos textos terem abreviações para as unidades, ou seja, *cm* para centímetros e *mm* para milímetros, enquanto que quando o texto é ditado a palavra é escrita sem abreviações. Sendo assim, como o *parser* usado está preparado para entender unidades escritas na forma abreviada, o tamanho das massas apenas é extraído nos textos colados ou escritos na parte da interface de inserção manual. Contudo é trivial mudar o *parser* para incorporar a palavra soletrada para cada tamanho da unidade ou outro tipo de unidades.

É de salientar que erros tipográficos como "estoma", que podem normalmente ocorrer nos texto escritos, não irão ocorrer nos textos ditados, o que traz uma vantagem às ferramentas *Speech-to-Text*.

A Figura 7.2 mostra o resultado retornado pela *Web Speech API* quando ditado para ela os relatórios da Figura 7.1

Como seria expectável neste tipo de ferramenta, ou seja, uma ferramenta que não tem qualquer tipo de aprendizagem do perfil de voz do utilizador e sendo o idioma usado o Português (idioma ainda pouco explorado no que toca a *Speech-To-Text*), o texto extraído tem algumas secções sem sentido. Contudo os principais termos BI-RADS[®], termos estes que são relevantes para o preenchimento do formulário, são capturados pela ferramenta a partir do texto ditado.

7.3 Discussão

No que concerne à experiência 1 podemos diretamente concluir que a *Web Speech API* é sensível ao tipo de voz, ou seja, é sensível à entoação e forma de falar do utilizador. Por exemplo a pessoa B foi a que levou a API à menor taxa de erro, sendo esta taxa

de 18,6% como pode ser visto na Tabela 7.2. Apesar da pessoa A ser a que possui a taxa mais alta de termos quase corretos, ou seja, 8,5% de termos que podem ser automaticamente corrigidos, a sua taxa de erro, 22,9% também é a maior.

Se usarmos a última coluna da Tabela 7.2 como medidor de performance da *Web Speech API* podemos concluir que a pessoa B seria a melhor para realizar a tarefa de ditar textos. Mais uma vez queremos lembrar que a *Web Speech API* não aprende nenhum tipo de padrão de voz, ou seja, não realiza o ajuste ao utilizador criando um perfil de voz para o mesmo no início do uso da ferramenta, ao contrário do constatado pela literatura sobre muitas das ferramentas pagas e com suporte à língua Inglesa.

Por outro lado, se medirmos a performance da API de acordo com o tipo de equipamento usado, claramente o uso de microfone externo é o que apresenta melhores resultados, seguido do *Smartphone* e por último o uso do portátil com microfone interno.

No que concerne à Experiência 2 é possível aprender duas lições. A primeira é que as ferramentas de *Speech to text* em português usadas para frases longas precisam de ser melhoradas. A *Web Speech API* é muito sensível à velocidade do discurso e não parece muito bem treinada para a língua em questão. A segunda lição que podemos retirar desta experiência é que apesar do texto reconhecido para as frases longas muitas vezes diferir do relatório original os termos BI-RADS[®] relevantes continuam a ser bem reconhecidos.

Capítulo 8

Conclusão

A interface das aplicações é bastante importante para que exista uma adoção completa das novas tecnologias nos sistemas de saúde. Neste trabalho exploramos o domínio do cancro da mama e produzimos para o utilizador uma interface onde os relatórios médicos podem ser ditados, em vez de utilizar formulários para inserir os dados clínicos.

Da realização deste trabalho consegue-se concluir que as ferramentas de reconhecimento de discurso com suporte a língua portuguesa de Portugal são muito reduzidas. Para além disso, o vocabulário extenso de uma ferramenta comum não é o mais adequado para o uso clínico. A qualidade do texto extraído ainda é muito fraca, porém somente alguns termos são necessários para algumas aplicações. Para o caso da extração de termos BI-RADS[®], os sistemas testados mostraram uma boa qualidade. As aplicações que suportam a língua portuguesa não garantem confidencialidade dos dados, o que, na área médica, é de extrema importância. Além disso, a qualidade depende da pessoa que realiza o discurso. O uso do Coruja em conjunto com o Julius resolve o problema da confidencialidade dos dados mantendo a mesma performance da *Web Speech API*. Todos os objetivos propostos nesta dissertação foram cumpridos, pois no final deste trabalho temos uma implementação completamente funcional de uma ferramenta de suporte à entrada de texto em Português, escrito e falado.

Assim sendo depreende-se que o próximo passo para esta ferramenta, apesar de complexo, deverá ser a elaboração de uma ferramenta própria para reconhecer termos em Português que seja independente da entoação de voz e que possa ser treinada apenas com o conjunto de palavras usado no âmbito do cancro da mama. Um outro caminho

que o projeto pode seguir será adaptar o MammoClass, ou então o *parser*, para que o formulário consiga ser totalmente preenchido de forma automática quando o texto é ditado ou digitado. Também seria interessante avaliar se uma caneta digital poderia ser incluída de forma a dotar o MammoClass de mais uma ferramenta de entrada de dados de forma a preencher os campos do formulário que não são extraídos pelo *parser*.

Bibliografia

- Adami, André Gustavo. 2010. Automatic speech recognition: From the beginning to the Portuguese language. *In: 9th International Conference on Computational Processing of the Portuguese Language*.
- Basma, S., Lord, B., Jacks, L. M., Risk, M., & M., Scaranelo A. 2011. Error rates in breast imaging reports: comparison of automatic speech recognition and dictation transcription. *AJR Am J Roentgenol*, **197**, 923–927.
- Bontrager, Kenneth. 2006. *Tratado de Técnica Radiológica e Base Anatômica*. Guanabara Koogan.
- Coulouris, George, Dollimore, Jean, Kindberg, Tim, & Blair, Gordon. 2011. *Distributed Systems: Concepts and Design*. 5th edn. USA: Addison-Wesley Publishing Company.
- Cunha, Filipe. 2011 (9). *Extracção de Atributos de Textos Clínicos Sobre Mamografias*. M.Phil. thesis, Faculdade de Ciências da Universidade do Porto.
- Dores, Liliana. 2014 (11). *Análise de dados de cancro da mama de duas populações geograficamente diferentes: Estados Unidos da América e Portugal*. M.Phil. thesis, Faculdade de Ciências da Universidade do Porto and Faculdade Medicina da Universidade do Porto.
- du Toit, Jacqueline, Hattingh, Retha, & Pitcher, Richard. 2015. The accuracy of radiology speech recognition reports in a multilingual South African teaching hospital. *BMC Medical Imaging*, **15**(1), 1–5.
- Ferreira, Pedro. 2010 (10). *Aplicação de Algoritmos de Aprendizagem Automática para a Previsão de Cancro de Mama*. M.Phil. thesis, Faculdade de Ciências da Universidade do Porto.

- Ferreira, Pedro, Fonseca, Nuno A, Dutra, Inês, Woods, Ryan, & Burnside, Elizabeth. 2015. Predicting malignancy from mammography findings and image-guided core biopsies. *International journal of data mining and bioinformatics*, **11**(3), 257–276.
- Figueiredo, F., & Dutra, I. 2016. *Otimização de Software de Reconhecimento de Discurso Aplicado à Oncologia*.
- Gold, Ben, Morgan, Nelson, & Ellis, Dan. 2011. *Speech and Audio Signal Processing Processing and Perception of Speech and Music*. 2 edn. Hoboken, NJ: Wiley-Interscience.
- Hammana, I., Lepanto, L., Poder, T., & Bellemare, C. Ly, M. S. 2015. Speech recognition in the radiology department: a systematic review. *HIM J.*, **44**(2), 4–10.
- Henricks, W. H., Roumina, K., & Skilton, B. E. 2002. The utility and cost effectiveness of voice recognition technology in surgical pathology. *Mod Pathol*, **15**.
- Hoyt, R., & Yoshihashi, A. 2010. Lessons Learned from Implementation of Voice Recognition for Documentation in the Military Electronic Health Record System. *Perspectives in Health Information Management*.
- Huang, Xuedong, Acero, Alex, & Hon, Hsiao-Wuen. 2001. *Spoken Language Processing: A Guide to Theory, Algorithm, and System Development*. 1st edn. Upper Saddle River, NJ, USA: Prentice Hall PTR.
- Kang, Hyunseok Peter, Sirintrapun, S. Joseph, Nestler, Rick J., & Parwani, Anil V. 2010. Experience With Voice Recognition in Surgical Pathology at a Large Academic Multi-Institutional Center. *American Journal of Clinical Pathology*, **133**(1), 156–159.
- Kopans, Daniel B. 2000. *Imagem da Mama*. 2 edn. Rio de Janeiro: MEDSI Editora Médica e Científica Ltda.
- Kurose, James F., & Ross, Keith. 2005. *Computer Networking: A Top-Down Approach Featuring the Internet*. 3rd edn. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc.
- McGurk, S., Brauer, K., Macfarlane, T. V., & Duncan, K. A. 2008. The effect of voice recognition software on comparative error rates in radiology reports. *The British Journal of Radiology*, **81**, 767–770.

- Nassif, H., Woods, R., Burnside, E., Ayvaci, M., Shavlik, J., & Page, D. 2009 (Dec). Information Extraction for Clinical Data Mining: A Mammography Case Study. *Pages 37–42 of: 2009 IEEE International Conference on Data Mining Workshops.*
- Nassif, Houssam, Cunha, Filipe, Moreira, Inês C., Cruz-Correia, Ricardo, Sousa, Eliana, Page, David, Burnside, Elizabeth S., & de Castro Dutra, Inês. 2012. Extracting BI-RADS features from Portuguese clinical texts. *Pages 1–4 of: 2012 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2012, Philadelphia, PA, USA, October 4-7, 2012.*
- Orvalho, Lurdes. 2008. *Cancro da mama: detectar é fundamental.*
- Patel, R., Greenberg, B., Montner, S., Funaki, A., Straus, C., Zangan, S., & MacMahon, H. 2013. *Reduction of Voice Recognition Errors in Radiological Dictation: Effects of Systematic Individual Feedback.*
- Pisco, João. 2001. *Noções Fundamentais de Imagiologia.* Lidel.
- Rabiner, Lawrence R., & Juang, B. H. 1993. *Fundamentals of speech recognition.* United states ed edn. PTR Prentice Hall.
- Reynolds, Douglas A., Quatieri, Thomas F., & Dunn, Robert B. 2000. Speaker Verification Using Adapted Gaussian Mixture Models. *Digit. Signal Process.*, **10**(1), 19–41.
- Rocha, Ricardo Daniel, Ferreira, Pedro Miguel, Dutra, Inês, Cruz-Correia, Ricardo, & Burnside, Elizabeth S. 2016. A Speech-to-Text Interface for MammoClass. *In: IEEE Computer-Based Medical Systems (CBMS).* Dublin and Belfast: IEEE, for IEEE.
- Siegel, Rebecca L., Miller, Kimberly D., & Jemal, Ahmedin. 2016. Cancer statistics, 2016. *CA: A Cancer Journal for Clinicians*, **66**(1), 7–30.
- Yen, Po-Yin, & Bakken, Suzanne. 2012. Review of health information technology usability study methodologies. *Journal of the American Medical Informatics Association*, **19**(3), 413–422.

Capítulo 9

Anexos

9.1 Termos BI-RADS[®] extraídos pela *Web Speech API*

As tabelas seguintes mostram todos os dados extraídos pela *Web Speech API* quando do ditado dos termos BI-RADS[®] pelos 4 indivíduos que participaram na experiência.

Testes Web Speech API

Termo	A – PC c/ Micro NGS	Confiança Observacoes	A – PC c/ Micro Integrado	Confiança Observacoes
arredondada	arredondada	0,85	arredondada	0,65
redonda	redonda	0,87	redonda	0,87
oval	oval	0,57	olival	0,18
ovóide	al fayed	0,65	ovóide	0,70
alongada	alongado	0,24 Género	alongado	0,22 Género
ovular	ova lar	0,68	alveolar	0,67
lobular	lobular	0,33	lobular	0,58
Polilobular	polly lobular	0,53 Inglês	polly lobular	0,48 Inglês
irregular	o que é regular	0,51	yahoo celular	0,67
circunscrita	circunscrita	0,92	circunscrita	0,92
regular	regular	0,77	regular	0,87
bem definida	bem definida	0,78	bem definida	0,83
bem delimitada	bem delimitada	0,87	bem delimitada	0,86
microlobular	microlopes polar	0,52	microalbumina	0,73
obscuras	obscuros	0,27 Género	obscuros	0,70 Género
obscorecida	auto escurecida	0,00	obscuros lda	0,23
Indistintas	índices tintas	0,67	índices tintas	0,84
Imprecisas	imprecisão as	0,32	intercidades	0,86
Indefinida	indefinida	0,92	indefinida	0,92
Espiculada	isp colada	0,58	isp colada	0,59
alta	alta	0,71	alta	0,70
elevada	elevada	0,84	elevada	0,84
hipertensa	hipertensa	0,20	hipertensa	0,64
igual	igual	0,90	igual	0,80
isodenso	esod anso	0,34	episódios hut	0,45
homogênea	homogênea	0,80	homogênea	0,75
baixa	baixa	0,73	baixa	0,69
ténue	terrenos	0,20	thenude	0,18
radiolucente	radiolucente	0,55	radiolucente	0,27
contém gordura	contem gordura	0,62	contem gordura	0,73
calcificações dérmicas	calcificações dérmicas	0,84	calcificações dérmicas	0,84
calcificações pele	calcificações pele	0,82	calcificações pele	0,55
calcificações vasculares	calcificações vasculares	0,92	calcificações vasculares	0,92
calcificações grosseiras	calcificações grosseiras	0,92	calcificações grosseiras	0,92
pipoca	pipoca	0,89	pipoca	0,92
calcificações em forma de bastonete	calcificações em forma de bastonete	0,83	calcificações em forma de bastonete	0,78
calcificações redondas	calcificações redondas	0,81	calcificações redondas	0,81
calcificações com centro lucente	calcificações com centro no sentido	0,56	calcificações com centro no siente	0,68
calcificações em casca de ovo	calcificações em casca de ovo	0,68	calcificações em casca de ovo	0,70
calcificações ductais ou leite calcio	calcificações tuc tais leite calcio	0,63	calcificações do que tais ou leite calcio	0,58
calcificações cicatriciais	calcificações cicatriciais	0,84	calcificações cicatriciais	0,93
citosteatonecrose	cic.fr teatro tuna croze	0,37	citou este ato necrose	0,74
calcificações distróficas	calcificações distróficas	0,83	calcificações distróficas	0,89
calcificações punctiformes	calcificações puntiformes	0,92 Acordo ortográfico	calcificações puntiformes	0,92 Acordo ortográfico
calcificações amórficas ou indistintas	calcificações amor ficas ou índice tintas	0,80	calcificações amor ficas ou índice tintas	0,72
dispersas morfológicamente	dispersas morfológicamente	0,83	dispersos morfológicamente	0,81 Género
calcificações pleomórficas ou heterogêneas	calcificações pleomórficas ou heterogêneas	0,82	calcificações pleomórficas ou heterogênea x	0,73 Número
polimórficas	polymorph caras	0,50	polymorph caras	0,53
calcificações finas ou lineares	calcificações finas ou lineares	0,77	calcificações finas ou lineares	0,81
agrupadas	agrupadores	0,18	agrupadores	0,50
vários núcleos	vários núcleos	0,92	vários núcleos	0,92
linear	limiar	0,53	limiar	0,72
area microcalcificações lineares	área microcalcificações lineares	0,75	área microcalcificações lineares	0,71

Testes Web Speech API

segmentar	sexo mentar	0,56		segmentar	0,92	
regional	regional	0,91		regional	0,92	
área extensa	área extensa	0,78		área extensa	0,85	
difusa	difusa	0,76		difusa	0,89	
dispersas morfológicamente	dispersas morfológicamente	0,82		dispersas morfológicamente	0,88	
ambos os lados	ambos os lados	0,92		ambos os lados	0,92	
distorção arquitectural	distorção arquitetural	0,92	Acordo ortográfico	distorção arquitectural	0,84	
desorganização arquitectural	desorganização architectural	0,76	Inglês	desorganização architectural	0,76	Inglês
ducto dilatado	ducto dilatado	0,78		ducto dilatado	0,78	
carcinoma ductal	carcinoma ductal	0,92		carcinoma ductal	0,90	
Gânglio linfático intramamário	gânglio linfático intramamario	0,69		gânglio linfático intramamario	0,69	
Gânglio mamário	gânglio mamario	0,59		gânglio mamario	0,60	
densidade assimétrica	densidade assimétrica	0,65		densidade assimétrica	0,77	
assimetria mamária	assimetria mamária	0,92		assimetria mamária	0,92	
densidade focal	densidade cal	0,69		densidade focal	0,81	
foco nodular	foco nodular	0,81		foco nodular	0,82	
foco assimétrico	foco assimétrico	0,77		foco assimétrico	0,73	
retracção cutânea	reparação cutânea	0,74		reparação cutânea	0,52	
repuxamento da pele	re tratamento da pele	0,90		repuxamento da pele	0,55	
retracção do mamilo	retração do mamilo	0,91	Acordo ortográfico	retração do mamilo	0,88	Acordo ortográfico
inversão mamilar	inversão mamilar	0,82		inversão mamilar	0,81	
espessamento cutâneo	espessamento cutâneo	0,73		espessamento cutâneo	0,72	
edema	edema	0,86		edema	0,92	
eritema	eritema	0,92		eritema	0,92	
mastite	mastite	0,76		mastite	0,82	
espessamento trabecular	espessamento trabecular	0,84		espessamento trabecular	0,79	
lesão cutânea	lesão cutânea	0,86		lesão cutânea	0,84	
ulceração cutânea	ulceração cutânea	0,82		ulceração cutânea	0,82	
adenopatia axilar	adenopatia axilar	0,92		adenopatia axilar	0,92	
adenomegalia axilar	adenomegalia axilar	0,00		adenomegalia axilar	0,89	
cavado axilar	cavado axilar	0,76		cavado axilar	0,75	
axila positiva	axila positiva	0,87		axila positiva	0,87	
Gânglio axilar suspeito	gânglio axilar suspeito	0,86		gânglio axilar suspeito	0,85	

Testes Web Speech API

A – Smartphone	Confiança	Observacoes	C – PC c/ Micro NGS	Confiança	Observacoes	C – PC c/ Micro Integrado
arredondada	0,63		arredondada	0,92		arredondada
redonda	0,95		redonda	0,91		redonda
oval	0,76		oval	0,33		oval
offside	0,30		ovóide	0,64		ovni
alongador	0,24		alongar a vida	0,30		alongar a vida
of al'ar	0,51		alvalade	0,65		ova online
lobular	0,66		lobular	0,43		lobular
polly lobular	0,57	Inglês	polly lobular	0,31	Inglês	polly lobular in
ir regular	0,71		ir regulares	0,65	Número	iris online
circunscrita	0,92		circunscrita	0,92		circunscrita
regular	0,86		regular	0,71		regular
bem definida	0,84		bem definida	0,76		bem definida
bem delimitada	0,79		bem delimitada	0,90		bem delimitada
microalbumina	0,42		me trollou volare	0,25		microglobulina
ovos curas	0,34		ovos curas	0,42		ovos curas
discursiva	0,32		ovos conhecida	0,51		obscura sida
índices tintas	0,50		inexistentes	0,30		índices tintas
imprecisões	0,00		imprecisões	0,00		indecisas
indefinida	0,92		indefinida	0,90		indefinida
isp colada	0,61		espiculada	0,00		espiculada
alta	0,69		alta	0,26		alta
elevada	0,92		elevada	0,76		elevada
império bonança	0,19		hipertensão	0,76		hipertensão
igual	0,92		igual	0,61		igual
episode 1 sub	0,28		isodense	0,92	Inglês	esod ançã
homogênea	0,79		homogênea	0,74		homogênea
baixa	0,92		baixa	0,68		baixa
tenor	0,50		penugem	0,00		bmw
radiolucente	0,55		radiolucente	0,56		radiolucente
contem cordura	0,35		contem gordura	0,48		contém gordura
calcificações dérmicas	0,92		calcificações dérmicas	0,92		calcificações dérmicas
calcificações bela	0,58		calcificações papel	0,59		calcificações papel
calcificações vasculares	0,92		calcificações vasculares	0,92		calcificações vasculares
calcificações grosseiras	0,92		calcificações grosseiras	0,92		calcificações grosseiras
pipoca	0,92		pipoca	0,87		pipoca
calcificações em forma de bastonete	0,81		calcificações em forma de bastonete	0,75		calcificações em forma de bastonete
calcificações redondas	0,92		calcificações redondas	0,81		calcificações redondas
calcificações com centro lucent	0,59	Inglês	calcificações com centro no sentido	0,57		calcificações com centro luciano
calcificações em casca do ovo	0,77	Género	calcificações em casca de ovo	0,84		calcificações em casca de ovo
calcificações ductais ou leite calcio	0,71		calcificações ductais leite calcio	0,68	Falta Palavra	calcificações do que tens ou leite calcio
calcificações cicatriciais	0,86		calcificações cicatriciais	0,92		calcificações cicatriciais
se tu és teatro necrose	0,83		situação esta é toni kross	0,26		citywest de toni kross
calcificações distróficas	0,86		calcificações distróficas	0,78		calcificações distróficas
calcificações puntiformes	0,92	Acordo ortográfico	calcificações puntiformes	0,92	Acordo ortográfico	calcificações puntiformes
calcificações amor ficas ou índice pintas	0,57		calcificações amor ficas ou instincts	0,62	Inglês	calcificações amor ficas ou índice tintas
dispersas morfológicamente	0,89		dispersas morfológicamente	0,69		dispersas morfológicamente
calcificações pleomórficas ou heterogênea	0,48	Número	calcificações pleomórficas ou heterogêneas	0,78		calcificações pleomórficas ou heterogênea
polymorph caras	0,08		polymorph caras	0,46		porn mov ficas
calcificações finas ou lineares	0,66		calcificações finas ou lineares	0,79		calcificações finas ou lineares
agrupadores	0,18		agrupar linhas	0,45		agrupar linhas
vários núcleos	0,92		vários núcleos	0,92		vários núcleos
linear	0,45		line	0,69		linear
área microcalcificações lineares	0,79		área microcalcificações lineares	0,58		área microcalcificações lineares

Testes Web Speech API

segmentar	0,92
regional	0,92
área extensa	0,63
difusa	0,91
dispersas morfológicamente	0,72
ambos os lados	0,92
distorção arquitectural	0,84
desorganização architectural	0,74 Inglês
ducto dilatado	0,80
carcinoma ductal	0,93
gânglio linfático intramamario	0,70
gânglio mamario	0,60
densidade assimétrica	0,61
assimetria mamária	0,92
densidade focal	0,89
foco nodular	0,89
foco assimétrico	0,77
reparação cutânea	0,66
repuxamento da pele	0,90
retração do mamilo	0,91 Acordo ortográfico
inversão mamilar	0,73
espessamento cutâneo	0,73
edema	0,81
eritema	0,93
mastite	0,87
espessamento trabecular	0,92
lesão cutânea	0,87
ulceração cutânea	0,81
adenopatia axilar	0,92
adenomegalia axilar	0,89
cavado axilar	0,76
axila positiva	0,84
gânglio axilar suspeito	0,92

segmentar	0,72
regional	0,86
área extensa	0,76
difusão	0,79
dispersas morfológicamente	0,77
ambos os lados	0,92
distorção arquitectural	0,92
desorganização architectural	0,81 Inglês
youtube dilatado	0,50
carcinoma ductal	0,91
gânglio linfático intramamario	0,70
gânglio mamario	0,58
densidade assimétrica	0,52
spm maia	0,49
densidade focal	0,47
foco nodular	0,80
foco assimétrico	0,50
respiração cutânea	0,79
repuxamento da pele	0,31
retração do mamilo	0,81 Acordo ortográfico
inversão mamilar	0,81
espessamento cutâneo	0,72
edema	0,55
eritema	0,92
mastite	0,86
espessamento trabecular	0,88
lesão cutânea	0,78
ulceração cutânea	0,84
adenopatia axilar	0,92
adenomegalia axilar	0,88
cavado axilar	0,37
axila positiva	0,42
gânglio axilar suspeito	0,00

segmentar	
regional	
área extensa	
difusão	
dispersas morfológicamente	
ambos os lados	
distorção arquitectural	
desorganização architectural	
ducto dilatado	
carcinoma ductal	
gânglio linfático intramamario	
gânglio mamario	
densidade é simétrica	
assimetria mamária	
densidade focal	
foco nodular	
foco assimétrico	
respiração cutânea	
repuxamento da pele	
retração do mamilo	
inversão mamilar	
espessamento cutâneo	
edema	
eritema	
mastite	
espessamento tradutor online	
lesão cutânea	
ulceração cutânea	
adenopatia axilar	
adenomegalia axilar	
cavado axilar	
axila positiva	
gânglio axilar suspeito	

Testes Web Speech API

Confiança	Observacoes	C – Smartphone	Confiança	Observacoes	D – PC c/ Micro NGS	Confiança	Observacoes
0,90		arredondada	0,87		arredondada	0,92	
0,92		redonda	0,89		redonda	0,85	
0,71		oval	0,71		oval	0,59	
0,68		ovóide	0,60		omoide	0,85	
0,42		grávida	0,69		alongado	0,19	Género
0,19		alvalade	0,29		alveolar	0,54	
0,53		lobular	0,33		lobular	0,67	
0,55		polly lobular	0,32	Inglês	paulo gular	0,71	
0,59		yahoo celular	0,77		isolar	0,53	
0,92		circunscrita	0,92		circunscrita	0,92	
0,32		regular	0,70		regular	0,85	
0,81		bem definida	0,81		bem definida	0,78	
0,81		bem delimitada	0,90		bem delimitada	0,84	
0,47		microlog online	0,46		micromolar	0,40	
0,65		ovos curas	0,59		escuras	0,25	
0,20		ovos conhecida	0,33		ovos conhecida	0,06	
0,46		índices tintas	0,71		indistinto as	0,43	Género
0,92		e precisas	0,68		precisas	0,87	
0,93		indefinida	0,69		indefinida	0,61	
0,23		espiculada	0,00		colada	0,42	
0,70		alta	0,54		alta	0,64	
0,92		elevada	0,92		elevada	0,87	
0,83		hyper de pertença	0,49		iped ansa	0,32	
0,76		igual	0,78		igual	0,87	
0,27		isodense	0,92	Inglês	esod ansioso	0,44	
0,77		homogênea	0,84		homogênea	0,70	
0,82		baixa	0,90		baixa	0,74	
0,29		terno	0,32		terno	0,38	
0,37		radiolucente	0,55		radiolucente	0,55	
0,76		contém gordura	0,75		contém gordura	0,57	
0,68		calcificações térmicas	0,87		calcificações térmicas	0,92	
0,59		calcificações pele	0,79		calcificações palace	0,61	
0,92		calcificações vasculares	0,92		calcificações vasculares	0,92	
0,92		calcificações grosseiras	0,92		calcificações grosseiras	0,92	
0,89		pipoca	0,92		pipoca	0,84	
0,71		calcificações em forma de bastonete	0,72		calcificações em forma de bastonete	0,79	
0,80		calcificações redondas	0,84		calcificações redondas	0,92	
0,49		calcificações com centro lucent	0,57	Inglês	calcificações com centro lucent	0,37	Inglês
0,88		calcificações em casca de ovo	0,86		calcificações em casca de ovo	0,73	
0,55		calcificações do que tais leite calcio	0,72		calcificações do que tais leite calcio	0,45	
0,89		calcificações cicatriciais	0,85		calcificações cicatriciais	0,91	
0,17		citosteatonecrose	0,93		citosteatonecrose	0,92	
0,89		calcificações distróficas	0,91		calcificações distróficas	0,88	
0,92	Acordo ortográfico	calcificações puntiformes	0,92	Acordo ortográfico	calcificações puntiformes	0,92	Acordo ortográfico
0,55		calcificações amor ficas ou distintas	0,77		calcificações amor ficas ou distintas	0,75	
0,73		dispersas morfologicamente	0,79		dispersas morfologicamente	0,91	
0,62	Número	calcificações promote ficas ou heterogenia	0,64		calcificações pleomórficas ou heterogenia as	0,54	Número
0,60		polimorfa ficas	0,35		paulimar sicas	0,52	
0,64		classificações finas ou lineares	0,84		calcificações finas ou lineares	0,86	
0,38		agrupando 10	0,10		agrupado as	0,13	Género
0,92		vários núcleos	0,92		vários núcleos	0,92	
0,31		linear	0,39		limiar	0,44	
0,72		área microcalcificações lineares	0,75		área microcalcificações lineares	0,85	

Testes Web Speech API

0,56	segmentar	0,76	segmentar	0,79
0,71	regional	0,92	regional	0,91
0,78	área extensa	0,81	área extensa	0,85
0,65	difusão	0,86	difusa	0,77
0,80	dispersas morfologicamente	0,89	dispersas morfologicamente	0,91
0,92	ambos os lados	0,92	ambos os lados	0,92
0,52	distorção arquitetural	0,83	distorção arquitetural	0,92
0,76 Inglês	desorganização arquitetural	0,75 Inglês	desorganização arquitetural	0,84
0,79	ducto dilatado	0,29	do sul dilatado	0,78
0,88	carcinoma ductal	0,91	carcinoma ductal	0,92
0,70	gânglio linfático intramamario	0,70	gânglio linfático intramamario	0,70
0,60	gânglio mamario	0,60	gânglio mamario	0,60
0,41	densidade assimétrica	0,51	densidade assimétrica	0,55
0,92	assimetria mamária	0,92	assimetria mamária	0,92
0,79	densidade focal	0,80	densidade do cal	0,76
0,80	foco nodular	0,83	foco nodular	0,81
0,25	foco assimétrico	0,70	foco assimétrico	0,79
0,59	respiração cutânea	0,92	reparação cutânea	0,80
0,42	repuxamento da pele	0,31	repuxamento da pele	0,81
0,76 Acordo ortográfico	retração do mamilo	0,90 Acordo ortográfico	reparação do mamilo	0,89
0,86	inversão mamilar	0,57	inversão mamilar	0,86
0,73	espessamento cutâneo	0,74	espessamento cutâneo	0,73
0,17	edema	0,89	ibirama	0,26
0,93	eritema	0,87	eritema	0,46
0,83	mastite	0,68	mastite	0,92
0,57	espessamento trabecular	0,81	espessamento trabecular	0,84
0,89	lesão cutânea	0,91	lesão cutânea	0,84
0,86	ulceração cutânea	0,84	ulceração cutânea	0,88
0,92	adenopatia axilar	0,92	adenopatia axilar	0,92
0,17	adenomegalia axilar	0,85	adenomegalia axilar	0,75
0,76	cavado axilar	0,37	cavado axilar	0,00
0,73	axila positiva	0,81	axila positiva	0,72
0,80	gânglio axilar suspeito	0,83	gânglio axilar suspeito	0,92

Testes Web Speech API

D – PC c/ Micro Integrado	Confianca	Observacoes	D – Smartphone	Confianca	Observacoes	B – PC c/ Micro Integrado	Confianca
arredondada	0,91		arredondada	0,87		arredondada	0,90
redonda	0,91		redonda	0,89		redonda	0,82
oval	0,71		oval	0,68		oval	0,72
obaid	0,39		all boys band	0,38		ovóide	0,70
alongar a vida	0,57		alongar a vida	0,76		alugada	0,81
alveolar	0,17		alda lara	0,73		ovar	0,26
lobular	0,68		lobular	0,70		lobular	0,60
polly lobular	0,56	Inglês	polly lobular	0,55	Inglês	polly lobular	0,47
yahoo lara	0,25		yahoo celular	0,80		regular	0,82
circunscrita	0,90		circunscrita	0,91		circunscrita	0,70
regular	0,73		regular	0,70		regular	0,74
bem definida	0,91		bem definida	0,76		bem definida	0,74
bem delimitada	0,76		bem delimitada	0,86		bem delimitada	0,88
micromolar	0,21		microlopes luar	0,50		micro lobular	0,90
ovos curas	0,65		obscura	0,78	Número	obscuras	0,56
of course vida	0,50		obscure vida	0,62		discursiva	0,29
e distintas	0,75		e destinos	0,58		índice tintas	0,67
e precisas	0,69		e precisas	0,70		imprecisões	0,39
indefinida	0,77		indefinida	0,70		indefinida	0,81
isp colada	0,51		espe colada	0,52		especulativa	0,64
alta	0,59		alta	0,70		alta	0,43
elevada	0,92		elevada	0,92		elevada	0,92
hyper criança	0,73		hyperv ançã	0,42		hipertensa	0,18
igual	0,76		igual	0,44		igual	0,51
episódio 11	0,44		esod ansul	0,50		episode 1 sub	0,49
homogênea	0,74		homogênea	0,76		homogênea	0,76
baixa	0,87		baixa	0,76		baixa	0,76
terreno	0,32		reno	0,20		terreno	0,17
radiolucente	0,29		rádio luciano	0,73		radiolucente	0,37
contem gordura	0,72		contem gordura	0,73		como tem gordura	0,75
calcificações benignas	0,84		calcificações benignas	0,84		calcificações de micas	0,78
calcificações bela	0,55		calcificações bela	0,59		calcificações pele	0,55
calcificações vasculares	0,92		calcificações vasculares	0,92		calcificações vasculares	0,92
calcificações grosseiras	0,92		calcificações grosseiras	0,92		calcificações grosseiras	0,80
pipoca	0,91		pipoca	0,93		pipoca	0,66
calcificações em forma de bastonete	0,81		calcificações em forma de bastonete	0,79		calcificações em forma de bastonete	0,71
calcificações redondas	0,92		calcificações redondas	0,92		calcificações redondas	0,78
calcificações com centro ocidente	0,50		calcificações com centro luso dente	0,66		calcificações concentro lucent	0,64
calcificações em casca de ovo	0,69		calcificações em casca de ovo	0,88		calcificações em casca de ovo	0,88
calcificações do cais ao leite calcio	0,57		calcificações do pai ao leite calcio	0,64		calcificações ductais ou leite calcio	0,50
calcificações cicatriciais	0,89		calcificações cicatriciais	0,90		calcificações cicatriciais	0,91
citoesteatonecrose	0,92		se tu és teto necrose	0,47		citoesteatonecrose	0,65
calcificações distróficas	0,91		calcificações distróficas	0,90		calcificações distróficas	0,85
calcificações puntiformes	0,92	Acordo ortográfico	calcificações puntiformes	0,92	Acordo ortográfico	calcificações puntiformes	0,59
calcificações a morte cas ou índice tintas	0,56		calcificações a morte cas ou distintas	0,44		calcificações amor ficas ou indistintas	0,51
dispersas morfológicamente	0,48		descobertas morfológicamente	0,85		dispersas morfológicamente	0,73
calcificações pleomórficas ou heterogêneas	0,86		calcificações pleomórficas ou heterogenia	0,47	Número	calcificações pleomórficas outro gêmeas	0,75
polymorph caras	0,20		cole mohr dicas	0,66		polymorph caras	0,55
calcificações finas ou lineares	0,88		calcificações finas ou lineares	0,79		calcificações finas ou lineares	0,74
agrupar linhas	0,51		agrupadores	0,00		agrupados	0,70
vários núcleos	0,92		vários núcleos	0,92		vários núcleos	0,77
limiar	0,18		limiar	0,46		linear	0,61
área microcalcificações lineares	0,79		área microcalcificações lineares	0,75		área microcalcificações lineares	0,66

Testes Web Speech API

segmentar	0,51
regional	0,83
área extensa	0,82
difusa	0,78
descobertas morfológicamente	0,84
ambos os lados	0,79
distorção arquitetural	0,92
desorganização arquitetural	0,84
do porto dilatado	0,75
carcinoma ductal	0,92
gânglio linfático intramamario	0,70
gânglio mamario	0,60
densidade assimétrica	0,60
assimetria mamária	0,92
densidade local	0,78
foco nodular	0,81
foco assimétrico	0,77
reparação cutânea	0,79
repuxamento da pele	0,76
retração do mamilo	0,87 Acordo ortográfico
inversão mamilar	0,82
espessamento cutâneo	0,73
eliana	0,36
eritema	0,79
mastite	0,69
espessamento trabecular	0,82
lesão cutânea	0,88
ulceração cutânea	0,83
adenopatia axilar	0,92
adenomegalia axilar	0,75
cavado axilar	0,74
axila positiva	0,80
gânglio axilar suspeito	0,83

fermentar	0,87
regional	0,92
área extensa	0,81
dificuldade	0,61
descobertas morfológicamente	0,86
ambos os lados	0,72
distorção arquitetural	0,92
desorganização arquitetural	0,43
ducto dilatado	0,85
carcinoma ductal	0,92
gânglio linfático intramamario	0,70
gânglio mamario	0,60
densidade assimétrica	0,61
assimetria mamária	0,92
densidade focal	0,82
foco nodular	0,81
foco assimétrico	0,72
reparação cutânea	0,75
repuxamento da pele	0,77
retração do mamilo	0,87 Acordo ortográfico
inversão mamilar	0,82
espessamento cutâneo	0,81
mediana	0,60
eritema	0,74
mastite	0,77
espessamento trabecular	0,76
lesão cutânea	0,85
ulceração cutânea	0,81
adenopatia axilar	0,92
adenomegalia axilar	0,75
cavado axilar	0,00
axila positiva	0,80
gânglio axilar suspeito	0,87

segmentar	0,69
regional	0,92
área extensa	0,35
difusa	0,81
dispersas morfológicamente	0,80
ambos os lados	0,88
distorção arquitetural	0,92
desorganização arquitetural	0,46
ducto dilatado	0,79
carcinoma ductal	0,84
gânglio linfático intra-mamário	0,92
gânglio mamário	0,63
densidade assimétrica	0,23
assimetria mamária	0,86
densidade focal	0,79
foco nodular	0,90
foco assimétrico	0,51
retração cutânea	0,72
repuxamento da pele	0,72
retração do mamilo	0,90
inversão mamilar	0,92
espaçamento cutâneo	0,38
edma	0,59
eritema	0,92
mastite	0,72
espessamento trabecular	0,77
lesão cutânea	0,69
ulceração cutânea	0,28
adenopatia axilar	0,92
adenomegalia axilar	0,83
cavado axilar	0,68
axila positiva	0,84
gânglio axilar suspeito	0,81

Testes Web Speech API

Observacoes	B – Smartphone	Confiança	Observacoes	B – PC c/ Micro NGS	Confiança	Observacoes
	arredondada	0,92		arredondada	0,71	
	redonda	0,82		redonda	0,69	
	oval	0,78		oval	0,76	
	ovóide	0,48		ovóide	0,7	
	prolongada	0,63		alugada	0,6	
	ovar	0,17		ovular	0,3	
	lobular	0,76		lobular	0,32	
Inglês	polly lobular	0,48	Inglês	polly lobular	0,33	Inglês
	regular	0,75		irregular	0,76	
	circunscrita	0,92		circunscrita	0,92	
	regular	0,92		regular	0,86	
	bem definida	0,72		bem definida	0,8	
	bem delimitada	0,88		bem delimitada	0,88	
Espaço	microglobulina	0,92		micromolar	0,7	
	obscuras	0,90		obscuras	0,9	
	obscuros lda	0,61		obscuros lda	0,52	
	índice pintas	0,39		índice tintas	0,58	
	impressivarts	0,70		impressive	0,69	
	indefinida	0,92		indefinida	0,33	
	especulativa	0,65		especulativa	0,74	
	alta	0,18		alta	0,71	
	elevada	0,76		elevada	0,76	
	hipertensão	0,29		hipertensa	0,22	
	igual	0,92		igual	0,8	
	Episódio 1 sub	0,29		episódios	0,76	
	homogênea	0,76		homogênea	0,68	
	baixa	0,77		baixa	0,71	
	terreno	0,63		terreno	0,25	
	radiolucente	0,57		radiolucente	0,57	
	como tem gordura	0,60		contem gordura	0,78	
	calcificações dérmicas	0,80		calcificações dérmicas	0,93	
	calcificações pele	0,46		calcificações pele	0,46	
	calcificações vasculares	0,92		calcificações vasculares	0,91	
	calcificações grosseiras	0,92		calcificações grosseiras	0,92	
	pipoca	0,82		pipoca	0,86	
	calcificações em forma de bastonete	0,79		calcificações em forma de bastonete	0,9	
	calcificações redondas	0,74		calcificações redondas	0,81	
	calcificações com centro lucent	0,74	Falta caracter	calcificações com centro lucent	0,3	Falta caracter
	calcificações em casca do ovo	0,85	de/do	calcificações em casca de ovo	0,88	
	calcificações ductais ou leite calcio	0,72		calcificações ductais ou leite calcio	0,68	
	calcificações cicatriciais	0,85		calcificações cicatriciais	0,86	
	citoeateonecrose	0,92		citoeateonecrose	0,92	
	calcificações distróficas	0,57		calcificações distróficas	0,73	
	calcificações puntiformes	0,63	Acordo ortográfico	calcificações puntiformes	0,81	Acordo ortográfico
	calcificações amor ficas ou índice	0,75		calcificações amor ficas ou distintas	0,8	
	dispersas morfológicamente	0,77		dispersas morfológicamente	0,8	
	calcificações pleomórficas ou heterogênea	0,52	Número	calcificações pleomórficas ou heterogêneas	0,82	
	polymorph caras	0,53		polymorphic	0,25	
	calcificações finas ou lineares	0,61		calcificações finas ou linear	0,71	Falta plural
	agrupados	0,92		agrupados	0,92	
	vários núcleos	0,84		vários núcleos	0,92	
	linear	0,32		linear	0,17	
	área microcalcificações lineares	0,41		área microcalcificações lineares	0,77	

Testes Web Speech API

	segmentar	0,79		segmentar	0,7
	regional	0,90		regional	0,69
	área extensa	0,78		área extensa	0,74
	difusa	0,68		difusa	0,6
	dispersas morfológicamente	0,68		dispersas morfológicamente	0,61
	ambos os lados	0,88		ambos os lados	0,92
	distorção arquitetural	0,92		distorção arquitetural	0,92
	desorganização arquitetural	0,84		desorganização arquitetural	0,84
	ducto dilatado	0,84		ducto dilatado	0,87
	carcinoma ductal	0,91		carcinoma ductal	0,92
	gânglio linfático intra-mamário	0,92		gânglio linfático intra-mamário	0,92
	gânglio mamário	0,77		gânglio mamário	0,7
	densidade assimétrica	0,65		densidade assimétrica	0,63
	assimetria mamária	0,92		assimetria mamária	0,92
	densidade focal	0,70		densidade focal	0,74
	foco nodular	0,87		foco nodular	0,87
	foco assimétrico	0,57		foco assimétrico	0,51
	reparação cutânea	0,65		retração cutânea	0,68
	repuxamento da pele	0,84		repuxamento da pele	0,89
Acordo ortográfico	retração do mamilo	0,90	Acordo ortográfico	retração do mamilo	0,79
	inversão mamilar	0,92		inversão mamilar	0,92
	espessamento cutâneo	0,46		espessamento cutâneo	0,44
	edema	0,72		edema	0,86
	eritema	0,92		eritema	0,92
	mastite	0,88		mastite	0,9
	espessamento trabecular	0,87		espessamento trabecular	0,87
	lesão cutânea	0,78		lesão cutânea	0,81
	ulceração cutânea	0,66		ulceração cutânea	0
	adenopatia axilar	0,92		adenopatia axilar	0,92
	adenomegalia axilar	0,80		adenomegalia axilar	0,23
	cavado axilar	0,80		cavado axilar	0,84
	axila positiva	0,83		axila positiva	0,9
	gânglio axilar suspeito	0,91		gânglio axilar suspeito	0,76
					Falta acento

9.2 Instalação do Julius e Coruja num servidor

Como visto na tese para converter o áudio foi usada a ferramenta *SoX* que pode ser instalada da forma tradicional.

Já no que toca à instalação do Julius e Coruja seguimos o tutorial que se segue.

- Baixei o Coruja para instalar o Julius que vem junto com ele nesse link: [Coruja_Linux.rar](#);
- Instalei as seguintes dependências do Julius:
 - Comandos (necessário ser root):
 - ...# aptitude install build-essential zlib1g-dev flex
 - ...# aptitude install libasound2-dev libesd0-dev libsndfile1-dev
- Instalei o Julius a partir do diretório julius-4.1.3 da seguinte forma:

1. Comandos:

- ..\$./configure
- ..\$ make clean
- ..\$ make
- ..\$ sudo make install

2. Check version:

```
leonardo@leonardo:~/coruja_jlapsapi$ julius -version
JuliusLib rev.4.1.3 (fast)

Engine specification:
- Base setup      : fast
- Supported LM    : DFA, N-gram, Word
- Extension       :
- Compiled by    : gcc -g -O2

Library configuration: version 4.1.3
- Audio input
  primary A/D-in driver  : oss (Open Sound System compatible)
  available drivers      : oss
  wavefile formats       : RAW and WAV only
  max. length of an input : 320000 samples, 150 words
- Language Model
  class N-gram support   : yes
  word id unit           : short (2 bytes)
- Acoustic Model
  multi-path treatment   : autodetect
- External library
  file decompression by  : zlib library
- Process handling
  fork on adinnet input  : no

Try '-help' for more information.
```

- Baixei os modelos acústicos no link: [coruja_jlapsapi1.5.rar](#);
- Extraí o arquivo (.rar) do passo anterior na home (/home/leonardo)

1. Conteúdo do diretório após extração:

- /home/leonardo/coruja_jlapsapi
- └─ dic.temp
- └─ julius.jconf
- └─ LaPSAM1.5.am.bin
- └─ LaPSAM1.5.tiedlist
- └─ LaPSLM1.5.lm.bin

- Criei um arquivo (**audiolist**) para listar os arquivos de áudio em **.wav** (é necessário que o .wav esteja de acordo com o parâmetro *samFreq* definido no .jconf e que este tenha apenas um channel, ou seja, mono - com o *ffmpeg* e *sox* dá pra fazer isso);
- Editei o arquivo **julius.jconf** conforme anexo;
- Entrei no diretório onde se encontra o julius.jconf e executei o seguinte comando:
 - ..\$ julius -C julius.jconf &> out
- Após um curto período de tempo (cerca de 10s), o terminal devolve o comando, então executei o seguinte comando para obter o resultado filtrado:
 - ..\$ cat out | grep sentence1: | sed "sentence1://g"

Pronto! A partir daí, se quiser que a entrada seja por microfone, vc deve entrar no .jconf e comentar (#) os parâmetros:

```
#-input rawfile
...
#-filelist audiolist
```

E retirar o comentário de

-input mic.