



Programa Interdisciplinar de Pós-Graduação em

# **Computação Aplicada**

**Mestrado Acadêmico**

Rodrigo Freiberger Rönnau

Modelo Baseado em Processamento de Dados  
Heterogêneos para Aplicações de Apoio Clínico

São Leopoldo, 2017



Rodrigo Freiburger Rönnau

**MODELO BASEADO EM PROCESSAMENTO DE DADOS  
HETEROGÊNEOS PARA APLICAÇÕES DE APOIO CLÍNICO**

Dissertação apresentada como requisito parcial  
para a obtenção do título de Mestre, pelo  
Programa de Pós-Graduação em Computação  
Aplicada da Universidade do Vale do Rio dos  
Sinos – UNISINOS

Orientador: Dr. Sandro José Rigo

São Leopoldo

2017

R773m Rönnau, Rodrigo Freibergger.  
Modelo baseado em processamento de dados heterogêneos para aplicações de apoio clínico / Rodrigo Freibergger Rönnau. – 2017.  
115 f. : il. (algumas color.) ; 30 cm.

Dissertação (mestrado) – Universidade do Vale do Rio dos Sinos, Programa Interdisciplinar de Pós-Graduação em Computação Aplicada, 2017.  
“Orientador: Prof. Dr. Sandro José Rigo”.

1. Informática na medicina. 2. Clínica médica – Exames. 3. Processamento eletrônico de dados. 4. Processamento de imagens auxiliado por computador. I. Título.

CDU 004:61

Dados Internacionais de Catalogação na Publicação (CIP)  
(Bibliotecária: Carla Maria Goulart de Moraes – CRB 10/1252)

Rodrigo Freiburger Rönnau

Modelo Baseado em Processamento de Dados Heterogêneos para Aplicações de Apoio Clínico.

Dissertação apresentada à Universidade do Vale do Rio dos Sinos – Unisinos, como requisito parcial para obtenção do título de Mestre em Computação Aplicada.

Aprovado em 06/12/2017

BANCA EXAMINADORA

---

Prof<sup>ª</sup>. Dr<sup>ª</sup>. Marta Rosecler Bez – FEEVALE

---

Prof. Dr. Jorge Luís Victória Barbosa – Unisinos

---

Prof. Dr. Sandro José Rigo – Unisinos (Orientador)

Visto e permitida a impressão  
São Leopoldo,

Prof. Dr. Rodrigo da Rosa Righi  
Coordenador PPG em Computação Aplicada



## **AGRADECIMENTOS**

Gostaria de agradecer a todos que, de alguma maneira, contribuíram para a realização desta pesquisa, em especial:

Aos meus pais, Angela e Roberto, pelos ensinamentos, suporte e incentivo.

À Karoline, que esteve ao meu lado durante todo o desenvolvimento do trabalho.

Ao professor Dr. Sandro, orientador e amigo, pelo auxílio, paciência e confiança.

Aos participantes do Grupo de Processamento e Análise de Imagens Médicas (PAIM), da Universidade FEEVALE.

À Nicole, pelas discussões e contribuições.

Aos demais colegas e amigos que auxiliaram nesse projeto.

À CAPES, pelo apoio a esta pesquisa.

Muito obrigado!



## RESUMO

O uso de sistemas computacionais no auxílio à prática clínica vem sendo amplamente estudado atualmente, buscando a avaliação de suas possibilidades na melhoria da qualidade do atendimento prestado aos pacientes. Dentre as aplicações com essa finalidade, podem-se destacar aquelas que atuam sobre laudos médicos ou sobre os exames por imagem, em atividades que realizam a extração, disponibilização e uso de características adquiridas através do processamento desses documentos. Entretanto, ao observar a literatura da área, percebe-se que há uma lacuna na utilização combinada das informações obtidas em cada tipo de processamento, ao mesmo tempo em que são indicadas possibilidades relevantes de criação de aplicações compartilhando e integrando estas informações. Outra lacuna identificada está relacionada à interoperabilidade dos dados e dos resultados obtidos entre os diferentes sistemas já existentes. Com o intuito de contribuir para a solução das questões apresentadas, neste trabalho é proposto um modelo, com estrutura modular e expansível, que viabiliza o emprego de diferentes formatos de entrada com o objetivo de prover, de forma integrada, informações de apoio ao médico ou especialista. Os dados extraídos são disponibilizados de maneira estruturada através de padrões reconhecidos, viabilizando a interoperabilidade entre os sistemas e o seu uso a partir de diferentes aplicações computacionais. Foram construídos dois protótipos, utilizando como base o modelo proposto. Para apresentar o funcionamento e os benefícios de uso do modelo, bem como possibilitar a sua avaliação, foram descritos cenários que demonstram seu emprego. Tanto o modelo como os protótipos foram apresentados a 12 profissionais da saúde e a 35 profissionais da computação. Os participantes preencheram um questionário de avaliação. Como resultado, 97,8% dos entrevistados indicaram que o modelo proposto é útil e 76,6% pretendem utilizá-lo e/ou divulgá-lo.

**Palavras-Chave:** Informática médica. Dados heterogêneos. Processamento de imagens. Processamento de linguagem natural.



## ABSTRACT

The use of computer systems to aid in the clinical practice has been widely studied, seeking an evaluation of its possibilities in improving the quality of the care provided to patients. Among the applications for this purpose, it is possible to highlight those that act on medical reports or medical images, in activities that perform the extraction, storage and use of characteristics acquired through the processing of these documents. However, observing the literature of the area, it is noticed that there is a gap in the combined use of information obtained in each type of processing, while indicating, at the same time, relevant possibilities of applications sharing and integrating this information. Another identified gap is related to the interoperability of the data and results obtained between different existing systems. In order to contribute to the solution of the presented questions, this work proposes a model, with a modular and expandable structure, which makes possible the use of different input formats with the objective of providing, in an integrated way, support information to the physician or specialist. The extracted data are made available in a structured manner through recognized standards, allowing the interoperability between the systems and their use from different computational applications. Two prototypes were constructed, using as basis the proposed model. Scenarios that demonstrate the use and benefits of the model have been described and used in its evaluation. Both the model and the prototypes were presented to 12 health professionals and 35 computer professionals. Participants completed an evaluation questionnaire. As result, 97.8% of respondents indicated that the proposed model is useful and 76.6% intend to use it and/or disseminate it.

**Keywords:** Medical informatics. Heterogeneous data. Image processing. Natural language processing.



## LISTA DE FIGURAS

Figura 1: Classificação da pesquisa.....	26
Figura 2: Etapas de PLN usualmente aplicadas à laudos para extração de informações. ....	30
Figura 3: Fragmento de laudo.....	30
Figura 4: Esquema genérico de sistemas CADe.....	33
Figura 5: Aplicação de técnicas de segmentação em imagem de TC do tórax. ....	34
Figura 6: Resultado das etapas de destacamento das lesões e detecção de candidatos. ....	35
Figura 7: Aplicação de filtro Gaussiano com máscara 3x3 em imagem de TC do tórax. ....	36
Figura 8: Exemplo de ontologia utilizada para representar o conteúdo de laudos médicos.....	37
Figura 9: Arquitetura do openEHR. ....	40
Figura 10: Exemplo de consulta utilizando AQL.....	41
Figura 11: Fluxo do envio e recebimento de uma mensagem HL7 v3.....	42
Figura 12: Fragmento de uma mensagem padrão HL7 utilizando XML ITS. ....	42
Figura 13: Exemplo da estrutura de um documento padrão CDA. ....	43
Figura 14: Exemplo de representação de um paciente em HL7 FHIR usando JSON.....	45
Figura 15: Exemplo de diferentes descritores para um mesmo conceito. ....	46
Figura 16: Estrutura geral de um arquivo no padrão DICOM.....	47
Figura 17: Conjunto de <i>data elements</i> em um arquivo no padrão DICOM. ....	47
Figura 18: Relações entre os tipos de dados e os padrões apresentados. ....	49
Figura 19: Componentes e agrupamentos do <i>framework</i> para CDSS. ....	52
Figura 20: Interface de uso para validar o <i>framework plug-and-play</i> . ....	53
Figura 21: Arquitetura do sistema que utiliza dados heterogêneos na predição de melanoma.	54
Figura 22: Fluxo de processamento da plataforma CER Hub.....	55
Figura 23: Interface de uso da busca multimodal com pesquisa e resultado.....	57
Figura 24: Fluxo de funcionamento da busca multimodal. ....	58
Figura 25: Fluxo de informações no ambiente METEOR.....	60
Figura 26: Fluxo da geração de CDA-XML a partir das imagens. ....	61
Figura 27: Estrutura geral do modelo proposto.....	65
Figura 28: Exemplos de elementos que podem existir em cada camada do modelo proposto.	66
Figura 29: Simplificação obtida através do uso das interfaces de entrada e de saída. ....	70
Figura 30: Representação em XML de um fragmento de laudo no padrão HL7 FHIR. ....	71
Figura 31: Ontologia correspondente ao fragmento de laudo apresentado na Figura 30.....	72
Figura 32: Elementos do modelo que foram desenvolvidos no primeiro protótipo.....	75

Figura 33: Exemplo de página do portal de acesso ao repositório de dados. ....	77
Figura 34: Recursos, atributos e relacionamentos do HL7 FHIR utilizados. ....	78
Figura 35: Ferramenta para extração de características em exames de TC do tórax.....	79
Figura 36: Trecho do XML retornado pelo <i>parser</i> PALAVRAS.....	80
Figura 37: Imagens de exames contidos na coleção <i>LungCT-Diagnosis</i> . ....	81
Figura 38: Retorno no formato XML da consulta ao recurso <i>Patient</i> que representa a Maria. ....	83
Figura 39: Detecção de pontos de calcificação em imagens de TC. ....	84
Figura 40: Laudo resultante do paciente João. ....	85
Figura 41: XML retornado na consulta ao recurso <i>DiagnosticReport</i> do paciente João.....	86
Figura 42: Elementos adicionados ao sistema na construção do segundo protótipo.....	88
Figura 43: Exemplo de laudo preliminar gerado pelo segundo protótipo. ....	90
Figura 44: Representação do <i>DiagnosticReport</i> com os sinais vitais do paciente João. ....	91
Figura 45: Agrupamento dos recursos com os sinais vitais vinculados ao paciente. ....	92
Figura 46: Resultado do laudo preliminar gerado para o paciente João.....	93
Figura 47: Composição do sistema atual (vermelho) e elementos adicionados (laranja). ....	97
Figura 48: Resultado do agrupamento das respostas equivalentes nos grupos. ....	103

## LISTA DE TABELAS

Tabela 1: Segmentos extraídos de laudos médicos. ....	29
Tabela 2: Exemplos de propriedades referentes às imagens em arquivos no padrão DICOM. ....	48
Tabela 3: Composição de parte da estrutura de dados utilizada no CER Hub. ....	56
Tabela 4: Classificação dos trabalhos relacionados quanto às suas características. ....	63
Tabela 5: Exemplos de dados e suas correspondências nos padrões openEHR e HL7 FHIR. ....	69
Tabela 6: Métodos disponíveis para uso no FHIRClient. ....	77
Tabela 7: Laudos criados para testes de PLN. ....	82
Tabela 8: Possíveis valores e resultados da etapa de conferência dos laudos. ....	87
Tabela 9: Métodos adicionados ao FHIRClient para o segundo protótipo. ....	88
Tabela 10: Características disponíveis, regras e frases correspondentes. ....	89
Tabela 11: Cenário atual de uso de aplicações de apoio que empregam dados heterogêneos. ....	98
Tabela 12: Opinião dos profissionais da saúde sobre a utilidade e divulgação do modelo. ....	99
Tabela 13: Avaliação da estrutura e das contribuições do modelo proposto. ....	100
Tabela 14: Opinião dos profissionais da computação sobre a utilidade do modelo. ....	101



## LISTA DE SIGLAS

ADL	<i>Archetype Definition Language</i>
AQL	<i>Archetype Query Language</i>
API	<i>Application program interface</i>
CAD	<i>Computer-aided detection and diagnosis</i>
CADe	<i>Computer-aided detection</i>
CADx	<i>Computer-aided diagnosis</i>
CDA	<i>Clinical Document Architecture</i>
CDS	<i>Clinical decision support</i>
CDSS	<i>Clinical decision support system</i>
CER	<i>Comparative effectiveness research</i>
DICOM	<i>Digital Imaging and Communications in Medicine</i>
DSS	<i>Decision support system</i>
DTI	<i>Diffusion tensor imaging</i>
EHR	<i>Electronic Health Record</i>
EMR	<i>Electronic Medical Record</i>
FHIR	<i>Fast Healthcare Interoperable Resources</i>
HL7	<i>Health Level 7</i>
HSV	<i>Hue Saturation Value</i>
i2b2	<i>Informatics for Integrating Biology and the Bedside</i>
JSON	<i>JavaScript Object Notation</i>
JPEG	<i>Joint Photographic Experts Group</i>
LBP	<i>Local Binary Patterns</i>
MRI	<i>Magnetic Resonance Image</i>
NCIT	<i>National Cancer Institute Thesaurus</i>
PACS	<i>Picture Archiving and Communication System</i>
PDI	Processamento de imagens
PHR	<i>Personal Health Record</i>
PLN	Processamento de linguagem natural
PNG	<i>Portable Network Graphics</i>
REST	<i>Representational State Transfer</i>
RDF	<i>Resource Description Framework</i>
RIM	<i>Reference Information Model</i>

ROI	<i>Region of interest</i>
SNOMED-CT	<i>Systematized Nomenclature of Medicine – Clinical Terms</i>
SQL	<i>Structured Query Language</i>
TCIA	<i>The Cancer Imaging Archive</i>
UML	<i>Unified Modeling Language</i>
UMLS	<i>Unified Medical Language System</i>
URL	<i>Uniform Resource Locator</i>
VR	<i>Value representation</i>
XML	<i>eXtensible Markup Language</i>

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	<b>21</b>
1.1 Motivação .....	23
1.2 Questão de pesquisa .....	25
1.3 Objetivos .....	25
1.4 Metodologia .....	26
1.5 Organização do documento .....	27
<b>2 FUNDAMENTAÇÃO TEÓRICA</b> .....	<b>29</b>
2.1 Processamento de textos.....	29
2.2 Processamento de imagens.....	33
2.3 Ontologias.....	37
2.4 Padrões de representação de dados clínicos .....	38
2.4.1 OpenEHR .....	39
2.4.2 HL7 .....	41
2.4.3 SNOMED-CT .....	45
2.4.4 DICOM.....	46
2.5 Considerações.....	48
<b>3 TRABALHOS RELACIONADOS</b> .....	<b>51</b>
3.1 A modular framework for clinical decision support systems: medical device plug-and-play is critical .....	51
3.2 Redesigning EHRs and Clinical Decision Support Systems for the Precision Medicine Era	53
3.3 CER Hub: An informatics platform for conducting comparative effectiveness research using multi-institutional, heterogeneous, electronic clinical data .....	55
3.4 Multimodal medical information retrieval with unsupervised rank fusion.....	57
3.5 METEOR: An Enterprise Health Informatics Environment to Support Evidence-based Medicine .....	59
3.6 GIDAC: a prototype for bioimages annotation and clinical data integration.....	60
3.7 Análise comparativa e lacunas de pesquisa .....	62
<b>4 MODELO PROPOSTO</b> .....	<b>65</b>
4.1 Entradas de dados .....	66
4.2 Extração de características.....	67
4.3 Repositório de características .....	68
4.4 Aplicações.....	72
4.5 Considerações.....	74
<b>5 DESENVOLVIMENTO DOS PROTÓTIPOS</b> .....	<b>75</b>
<b>5.1 Conferência de laudos</b> .....	<b>75</b>
5.1.1 Repositório de características .....	76
5.1.2 Processamento de imagens médicas .....	78
5.1.3 Processamento de Linguagem Natural .....	80
5.1.4 Aplicação “Conferência de laudos” .....	81
5.1.5 Coleta de dados .....	81
5.1.6 Cenário de uso .....	83
<b>5.2 Geração de laudo preliminar</b> .....	<b>87</b>
5.2.1 Repositório de características .....	88
5.2.2 Processamento de dados de sensores .....	89
5.2.3 Aplicação “Geração de laudo preliminar”.....	89
5.2.4 Coleta de dados .....	90
5.2.5 Cenário de uso .....	91
<b>5.3 Considerações</b> .....	<b>93</b>
<b>6 AVALIAÇÃO</b> .....	<b>95</b>
<b>6.1 Cenários de uso</b> .....	<b>95</b>
6.1.1 Substituição/atualização do padrão de representação .....	95
6.1.2 Ampliação de funcionalidades.....	96
<b>6.2 Questionários</b> .....	<b>97</b>

6.2.1 Profissionais da saúde .....	98
6.2.2 Profissionais da computação .....	99
<b>6.3 Considerações.....</b>	<b>101</b>
<b>7 CONCLUSÃO .....</b>	<b>105</b>
<b>REFERÊNCIAS.....</b>	<b>109</b>

## 1 INTRODUÇÃO

No contexto clínico, os sistemas médicos de informação possuem um papel fundamental, auxiliando os profissionais no atendimento prestado aos pacientes e contribuindo para a obtenção de avanços importantes, de qualidade e eficiência, na prestação desses serviços (FICHMAN; KOHLI; KRISHNAN, 2011). Buntin et al. (2011) apresentam um trabalho de revisão neste contexto, no qual indicam que 92% das publicações relacionadas apresentam resultados positivos quanto ao uso de aplicações computacionais na área médica, trazendo benefícios para as pessoas, tanto em nível individual como em nível populacional. Como exemplo, Fichman, Kohli e Krishnan (2011) citam um software que avalia dados clínicos de determinada região, permitindo a identificação de eventos adversos que podem causar doenças contagiosas ou complicações para a saúde das pessoas.

O ramo clínico é amplamente diversificado, englobando uma vasta gama de profissionais (médicos, enfermeiros, administradores, entre outros) e um elevado número de partes interessadas (pacientes, fornecedores de serviços, órgãos reguladores, entre outros). As pesquisas neste domínio abrangem diferentes áreas de conhecimento, tais como sociologia, administração, tecnologia da informação, biologia, direito, entre outras (FICHMAN; KOHLI; KRISHNAN, 2011). Segundo Buntin et al. (2011) e Roehrs et al. (2017), as esferas que usualmente utilizam tecnologias da informação neste contexto estão relacionadas, principalmente, com o *Electronic Health Record* (EHR), sistemas de apoio à decisão clínica, intercâmbio de informações médicas, prescrição para pacientes ambulatoriais, *Personal Health Record* (PHR), controle de pacientes, telemedicina ou monitoramento remoto, recuperação de informações e funções administrativas.

Conforme descrito por Lobach et al. (2012), um sistema de apoio à decisão clínica é qualquer aplicação eletrônica utilizada para auxiliar o especialista na tomada de decisão médica. Tais aplicações consideram as características do indivíduo e geram opiniões ou recomendações, as quais podem ser apresentadas e utilizadas por profissionais para a construção do diagnóstico. Para representar esse tipo de sistema são utilizadas diferentes siglas, como CDS (*Clinical Decision Support*), CDSS (*Clinical Decision Support System*) ou simplesmente DSS (*Decision Support System*). Usualmente são empregados algoritmos que geram sugestões específicas para cada paciente, fazendo uso de informações da pessoa em relação a uma base de conhecimento (GARG et al., 2005).

Sistemas de diagnóstico assistido por computador (definidos pela sigla CAD, do inglês *computer-aided detection and diagnosis*) representam as aplicações que tem, como finalidade, auxiliar os profissionais na formulação do diagnóstico clínico. Estas aplicações fornecem aos médicos informações complementares sobre características do paciente detectadas e/ou classificadas, bem como dados adicionais que podem auxiliar o especialista na tomada de decisão. O resultado fornecido por estas aplicações é considerado uma segunda opinião sobre as condições do paciente, podendo explicitar associações não realizadas pelo profissional (SUZUKI, 2012; CHENG et al., 2016; MORRISON et al., 2016).

Segundo Firmino et al. (2016), os sistemas CAD são divididos em dois grandes grupos. O primeiro deles, representado pela sigla CADE (do inglês, *Computer-Aided Detection*) agrupa as aplicações cuja finalidade é detectar e delimitar lesões em exames médicos por imagem. O segundo grupo é identificado pela sigla CADx (do inglês, *Computer-Aided Diagnosis*), sendo utilizado na caracterização das lesões. Cheng et al. (2016) indicam que uma das principais aplicações de CADx está relacionada à definição de malignidade de

um nódulo, ou seja, determinar se a lesão representa um tumor maligno ou benigno, a partir da aplicação de técnicas de processamento de imagens nos exames.

Os sistemas de apoio ao diagnóstico podem ser desenvolvidos com diferentes finalidades, tais como: consolidação de informações dos exames e informações históricas dos pacientes (DEMNER-FUSHMAN; CHAPMAN; MCDONALD, 2009); redução do volume de dados a ser avaliado pelo especialista; identificação de características relevantes para o processo (EL-DAHSHAN et al., 2014); execução rápida e precisa de tarefas minuciosas, que demandam tempo, dedicação e atenção do responsável (KOCHENBORGER, 2013); apresentação de recomendações, entre outras atividades. De modo geral, os ganhos obtidos ao serem utilizados tais sistemas estão relacionados à redução do tempo necessário para obter o diagnóstico e à redução do número de falhas (CHENG et al., 2016; EL-DAHSHAN et al., 2014; SUZUKI, 2012; BOZKURT et al., 2016).

Exames médicos por imagem, como a tomografia computadorizada (TC), são frequentemente utilizados pelos profissionais da saúde pois permitem a avaliação das condições internas dos pacientes sem a necessidade de intervenções invasivas, como as cirurgias exploratórias. Segundo Firmino et al. (2016), a alta carga de trabalho dos radiologistas, associado ao elevado número de figuras que compõem um exame desse tipo, pode resultar em falhas na detecção das características ou de interpretação. Dessa forma, o uso de sistemas computacionais se tornou desejável para assistir o especialista na tomada de decisão.

O mesmo ocorre em relação aos textos médicos, nos quais o volume de dados é elevado. Neste caso, o uso de aplicações computacionais pode atuar como alternativa para um processo mais eficaz e adequado à área. Conforme Topaz et al. (2016), cerca de 80% desse tipo de informação está armazenada de maneira não estruturada. Essa condição limita o emprego de sistemas computacionais (PONS et al., 2016). Com relação à esta temática, conforme Bozkurt et al. (2016), a não popularização de DSS está relacionada ao trabalho manual existente atualmente para atender a necessidade de submeter, em diferentes sistemas, um mesmo conjunto de informações. Um exemplo é a inserção em um sistema especialista dos registros correspondentes a um laudo por profissionais que, usualmente, já o redigiram em outro momento, o que eleva as chances de que sugestões fornecidas pela aplicação sejam as mesmas que já foram inseridas no texto. Essa atividade consome tempo e é propensa a erros, já que os dados podem ser informados de forma incorreta ou incompleta.

Além da precisão, otimização do tempo e ganho de desempenho, uma característica é essencial para a popularização do uso de sistemas de apoio ao diagnóstico: a aplicação deve ser integrada ao fluxo de trabalho (van GINNEKEN; SCHAEFER-PROKOP; PROKOP, 2011). Isso significa que a execução das ferramentas de apoio e, conseqüentemente, a disponibilização dos seus resultados deve estar incorporada aos sistemas que os profissionais já utilizam atualmente. A partir dessa afirmação, em conjunto com as informações citadas anteriormente, pode-se concluir que um sistema de apoio ao diagnóstico deve possuir: processos eficientes e confiáveis de extração de características; uma maneira estruturada para armazenar as informações obtidas, permitindo a utilização posterior por sistemas computacionais; e, por fim, aplicações que sejam parte do fluxo de trabalho do usuário final.

## 1.1 Motivação

Diante do que foi apresentado na seção anterior, verifica-se que existem diferentes tipos de sistemas de apoio clínico. As diversas possibilidades relacionadas ao uso e desenvolvimento dessas aplicações (CSD, CDSS, DSS, CAD, CADe, CADx, entre outras) abrangem temas amplamente explorados atualmente, com o estado-da-arte em evolução constante, gerando assim evidente contribuição para os cuidados com a saúde das pessoas.

O estudo de trabalhos científicos e de sistemas disponíveis nesta área demonstra que, em geral, apenas um tipo de dado é utilizado em cada uma destas aplicações. Podem ser encontrados estudos que tratam de aplicações baseadas exclusivamente no processamento de textos (BOZKURT et al., 2016; PONS et al., 2016; TOPAZ et al., 2016). Outros trabalhos são fundamentados somente em processamento de imagens (CHENG et al., 2016; EL-DAHSHAN et al., 2014; FIRMINO et al., 2016). Desta forma, a análise das pesquisas atuais permite observar que não é comumente abordada a agregação de diferentes tipos de informações durante as atividades de apoio à prática médica. Há uma lacuna em relação à exploração e consolidação desses dados para compor um diagnóstico mais abrangente e preciso. Este aspecto será explorado neste trabalho, levando em conta os benefícios que podem ser obtidos com a integração dessas informações.

Segundo Riegler et al. (2016), as ferramentas CAD existentes, que utilizam técnicas de visão computacional, não fazem uso efetivo das múltiplas fontes de informações adicionais, tais como sensores, dados temporais e pareceres dos usuários. Especificamente sobre CADe e CADx, Firmino et al. (2016) afirmam que as áreas de detecção e diagnóstico relacionadas ao câncer são exploradas separadamente. Dessa forma, as aplicações cuja finalidade é prover diagnóstico não detectam nódulos e não tem bons níveis de automação, já as aplicações utilizadas na detecção das lesões não demonstram as características radiológicas dos tumores. Já no ambiente de análise de exames radiológicos, a demanda por algoritmos que fazem uso de fontes de dados heterogêneos é crescente, uma vez que diversos formatos de imagens são avaliados em busca de uma grande variedade de doenças e características, que podem resultar em um elevado número de possibilidades de diagnóstico. Considerando os diferentes cenários citados, é possível indicar que a combinação das informações prévias com consultas integradas a bases de conhecimento pode tornar o processo mais eficiente (WEESE; LORENZ, 2016).

Dentre os diferentes estudos que abordam o uso de dados heterogêneos em sistemas de apoio clínico, pode ser citado o de Morrison et al. (2016), em que foi proposta uma aplicação capaz de detectar observações que estão descritas em um laudo médico, determinar relações entre as observações identificadas e possíveis doenças, integrar esses dados com o histórico do paciente e gerar, como saída, um diagnóstico diferencial, que pode ser exibido para o profissional para auxiliar na tomada de decisão. Zhao et al. (2017) extraíram informações textuais em diferentes formatos contidos em EHRs, reunindo descrições dos diagnósticos, resultados de testes laboratoriais, relatórios de medicamentos utilizados, mantendo a sequência temporal das características obtidas, permitindo assim melhorias nos processos de classificação e identificação de padrões baseados em aprendizado de máquina. Em outra pesquisa, características extraídas previamente a partir de áudios, vídeos e textos foram empregadas, em algoritmos de classificação, com o objetivo de detectar indivíduos com depressão, demonstrando assim a necessidade de uso de informações de múltiplas fontes (PAMPOUCHIDOU et al., 2016). A fusão de características extraídas de dois diferentes tipos de exames por imagens, ressonância magnética (MRI) e imagem de difusão por ressonância

magnética (DTI), permitiu a classificação e identificação automática de casos de síndrome de Tourette, indicando que biomarcadores distintos podem ser empregados de maneira complementar no processo de diagnóstico (WEN et al., 2017).

Acredita-se que seja possível unir os resultados obtidos por trabalhos que atuam sobre cada tipo de dado, permitindo o uso de suas características de maneira complementar. Nesse sentido, podem ser conquistadas melhorias de precisão do conjunto de informações e pode ser viabilizado o desenvolvimento de novas ferramentas. Segundo Maglogiannis et al. (2015), sistemas de suporte clínico mais sofisticados se baseiam em registros de tratamentos prévios e outros dados clínicos, para prover recomendações específicas para os pacientes, reduzindo assim as falhas e aprimorando o atendimento. Considera-se que esta modalidade de aplicação possa aumentar a qualidade do processo de interpretação de exames, uma vez que as ferramentas automatizadas são frequentemente apoiadas em regras definidas por especialistas, que às constroem aplicando os conhecimentos mais recentes (FRIEDMAN; JOHNSON, 2006). Kansagra et al. (2015) afirmaram que os profissionais precisam avaliar os casos de maneira mais ampla, passando a considerar as características específicas dos pacientes, como seus dados laboratoriais, o histórico familiar, os exames físicos, além das imagens e suas respectivas observações, para que seja possível diagnosticar as condições com precisão e recomendar o tratamento adequado.

Conforme afirmado por Hazlehurst et al. (2015), para criar uma visão completa do estado de saúde do paciente são necessários diferentes tipos de dados, que podem estar armazenados de maneira estruturada ou não estruturada, em sistemas de controle dos hospitais, de cobrança de serviços clínicos, de laboratórios, de exames por imagem, além do prontuário do paciente. O uso de fontes complementares de informações, cujas características são extraídas e organizadas de forma automatizada, pode garantir que esse requisito seja atendido, uma vez que podem ser utilizados como entrada todos os dados disponíveis sobre o indivíduo.

Existem fontes de informações distintas na área médica, tais como textos (laudos, relatórios, bases de conhecimento, prontuários), imagens (exames médicos, atlas anatômicos), vídeos (exames médicos, animações que demonstram o funcionamento normal de estruturas), sons (exames médicos, gravações das observações dos especialistas) e informações obtidas a partir de sensores. Entende-se que seja possível utilizar esses dados heterogêneos de maneira integrada e que as aplicações resultantes poderão fornecer informações de maior relevância e precisão, se comparadas aos sistemas que são baseados em apenas um tipo de dado, contribuindo assim para o desenvolvimento das mesmas, ampliando a faixa de apoio da computação na área clínica e possibilitando a evolução do estado-da-arte.

A partir das lacunas observadas nos trabalhos estudados, visualiza-se a necessidade do desenvolvimento de um modelo que possa ser utilizado como base para a criação e emprego de aplicações clínicas com diferentes finalidades, como apoio ao diagnóstico, controle de pacientes, pesquisas e sistemas administrativos. Os componentes desse modelo devem ser modulares e de simples expansão, permitindo agregar funcionalidades e pesquisas já desenvolvidas sem que seja necessário modificar todo o sistema. As características obtidas devem ser armazenadas em uma base de dados estruturada e padronizada, possibilitando o seu uso por outras aplicações.

No âmbito da computação, a principal contribuição científica obtida nesta pesquisa é a definição dos elementos e relacionamentos do modelo proposto, que busca constituir uma estrutura de referência que auxilia os profissionais na construção de sistemas de apoio clínico que fazem uso de informações que estão em diferentes formatos e que são provenientes de

diferentes fontes (dados heterogêneos). Também visa simplificar a expansão de aplicações existentes com o uso de pesquisas já desenvolvidas, tornando-as mais completas e de maior abrangência.

## 1.2 Questão de pesquisa

A partir do cenário exposto anteriormente, a presente pesquisa pretende atender a seguinte questão:

Qual deve ser a composição e a organização de uma estrutura de referência para o desenvolvimento de aplicações de apoio clínico que fazem uso, de maneira integrada, de informações extraídas em diferentes tipos de dados e que viabilize a interoperabilidade com sistemas já desenvolvidos?

## 1.3 Objetivos

O objetivo geral deste trabalho é apresentar um modelo de sistema computacional que possa ser empregado como base para a construção de aplicações de apoio clínico com diferentes finalidades, possibilitando a utilização integrada de características extraídas de diferentes tipos de dados a partir do uso e armazenamento dos elementos obtidos de maneira estruturada.

Para permitir concretizar o objetivo geral será necessário concluir os seguintes objetivos específicos:

- a) Identificar o estado-da-arte em sistemas de apoio médico que utilizam informações extraídas de diferentes fontes, a partir de pesquisas publicadas em revistas científicas reconhecidas e outros repositórios digitais conceituados;
- b) Estabelecer uma solução satisfatória para a questão de pesquisa considerando o estado-da-arte e o contexto estudado;
- c) Propor um modelo capaz de extrair, armazenar, relacionar e manipular diferentes tipos de dados médicos;
- d) Desenvolver protótipos que exemplifiquem a utilização e a avaliação do modelo;
- e) Demonstrar a contribuição do uso do modelo proposto no apoio à medicina e à computação, avaliando os resultados gerados pelos protótipos desenvolvidos e os resultados obtidos nos questionários respondidos pelos profissionais;
- f) Propor trabalhos futuros relacionados ao emprego e melhorias do modelo apresentado.

A seção a seguir detalha a metodologia utilizada no desenvolvimento deste trabalho.

## 1.4 Metodologia

Do ponto de vista de sua natureza, o presente trabalho é caracterizado como uma pesquisa aplicada, já que a criação de protótipos com base no modelo proposto possibilitará demonstrar o seu uso prático, com o objetivo de solucionar um problema específico (PRODANOV; FREITAS, 2013).

É um trabalho com objetivos exploratórios, sendo sua finalidade melhorar o processo clínico a partir da integração de informações extraídas de diferentes fontes de dados. O desenvolvimento da pesquisa envolve buscas por informações sobre o tema, através de levantamento de bibliografias e análise de exemplos. A Figura 1 exibe os conceitos aplicados à metodologia deste trabalho, em negrito.

Figura 1: Classificação da pesquisa.



Fonte: Adaptado de Bez (2011).

Em relação aos procedimentos técnicos, são utilizados dois delineamentos. O primeiro deles se refere à utilização de materiais já publicados, como livros e artigos científicos referentes ao tema, caracterizando-a como uma pesquisa bibliográfica (PRODANOV; FREITAS, 2013). O segundo faz referência à validação com testes práticos, diferentemente da validação puramente por métodos teóricos. Dessa forma, o presente trabalho é também uma pesquisa experimental, já que a solução proposta envolve o desenvolvimento prático de protótipos de software, que serão submetidos a testes e avaliações (MARCONI; LAKATOS, 2003).

Por fim, a abordagem do problema é classificada como qualitativa, uma vez que as percepções de uso do modelo serão classificadas e analisadas, buscando avaliar a aplicação do método proposto para responder à questão de pesquisa. A avaliação será feita a partir dos resultados apresentados pelos cenários de uso das aplicações que serão desenvolvidas, que exploram as características do modelo proposto, e pela coleta de opiniões de profissionais e especialistas sobre os benefícios de uso do modelo em relação às diferentes abordagens atualmente utilizadas.

O método de trabalho a ser utilizado no desenvolvimento da pesquisa é formado por três etapas (WAZLAWICK, 2009), que são apresentadas e detalhadas a seguir:

- a) Descrever o modelo de extração, armazenamento e integração de dados médicos. O modelo proposto é composto por um conjunto de elementos, com diferentes finalidades, que estão interligados. A sua composição foi definida com base no estudo dos trabalhos relacionados e da bibliografia que aborda o tema de pesquisa;
- b) Desenvolver protótipos para avaliação do modelo. Os protótipos desenvolvidos são formados por elementos que existem em todas as camadas do modelo. É dada preferência pela utilização de métodos já desenvolvidos, demonstrando as possibilidades de integração e adaptação de cada componente. No desenvolvimento de todos os protótipos foi utilizada a linguagem de programação Java, a biblioteca HAPI FHIR para o componente que corresponde ao repositório de características, e diferentes técnicas de processamento de imagens e de processamento de linguagem natural, que foram detalhadas em outros trabalhos;
- c) Avaliar o comportamento do modelo proposto a partir da análise dos resultados gerados pelos protótipos em relação à diferentes cenários de uso. Apresentar o modelo e os protótipos a um grupo de 12 profissionais da saúde e a 35 profissionais da computação, com a finalidade de obter pareceres sobre as contribuições e benefícios de uso do modelo descrito.

## 1.5 Organização do documento

Este trabalho está organizado em sete capítulos. O Capítulo 2 apresenta a fundamentação teórica sobre os conceitos necessários para a elaboração do modelo. No Capítulo 3 são descritas, de maneira sucinta, pesquisas relacionadas ao uso de sistemas computacionais na prática médica que fazem uso de múltiplas fontes de dados. O Capítulo 4 detalha o modelo proposto, contendo informações sobre cada um dos seus elementos. No Capítulo 5 constam as características dos protótipos desenvolvidos com base no modelo proposto. O Capítulo 6 apresenta a avaliação do modelo, feita a partir de cenários de uso e com a participação de profissionais da saúde e da computação. O último Capítulo contém as considerações finais, contribuições, limitações e trabalhos futuros.



## 2 FUNDAMENTAÇÃO TEÓRICA

A construção de um sistema de apoio à prática médica pode envolver diferentes áreas da computação, dependendo das suas especificações. Nas subseções seguintes são apresentadas algumas das técnicas comumente utilizadas para extrair características de determinados tipos de dados, tais como os textos e as imagens. Por fim, são especificadas diferentes formas de representar as informações obtidas a partir do processamento das entradas, de maneira estruturada, permitindo a sua utilização posterior por aplicações computacionais.

### 2.1 Processamento de textos

De acordo com Liu e Rubin (2012), os EHRs são uma vasta fonte de informações clínicas na forma de anotações de evolução dos quadros dos pacientes, notas de médicos e especialistas, laudos de exames, resultados laboratoriais e relatórios de alta. Entretanto, esse material está, na maior parte dos casos, armazenado e descrito de maneira não estruturada, na forma de textos livres redigidos em linguagem natural, limitando assim a utilização de computadores em processos automáticos de acesso e análise dessas informações (RUBIN et al., 2007). Segundo Friedman e Johnson (2006), os textos escritos em linguagem natural são complexos de serem processados, de forma automática, por um computador, devido à grande variedade de expressões, palavras, termos e estruturas gramaticais que podem ser utilizadas para denotar, frequentemente, um mesmo conceito.

Segundo Agah (2013), o processamento de linguagem natural (PLN) é a análise automatizada da linguagem natural. O uso dessas técnicas visa contribuir para o atendimento da necessidade citada anteriormente, permitindo que as informações médicas que estão baseadas em textos sejam extraídas e representadas de maneira estruturada (FRIEDMAN; HRIPCSAK, 1999).

Em relatórios médicos, a execução de buscas por termos de forma isolada pode não ser o suficiente para que sejam alcançados os resultados esperados. Na Tabela 1 são apresentados três segmentos extraídos de diferentes laudos, que representam informações distintas com o uso de palavras semelhantes.

**Tabela 1: Segmentos extraídos de laudos médicos.**

<b>Laudo</b>	<b>Segmento de texto</b>
1	Ausência de derrame pleural.
2	Moderado derrame pleural bilateral.
3	Leve aumento do derrame pleural à direita

Fonte: Elaborado pelo autor.

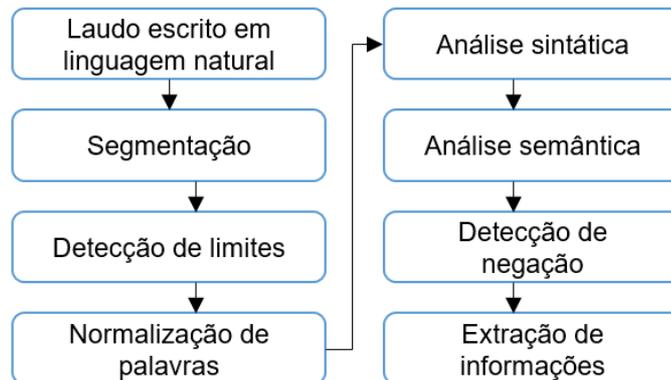
Se o objetivo for identificar os exames onde foi destacada a existência de derrame pleural, não basta pesquisar pelas palavras “derrame pleural”, uma vez o retorno dessa busca seria composto pelos laudos 1, 2 e 3. Entretanto, o laudo 1 indica justamente a ausência desta condição. No exemplo apresentado é importante entender a semântica das frases que formam

o laudo, já que um mesmo termo recebe diferentes significados dependendo das demais palavras que formam cada sentença.

Além disso, outra necessidade que recebe destaque, principalmente relacionada a textos que descrevem os achados em imagens médicas, é a identificação de valores numéricos, que fazem referência às observações apresentadas pelo médico. É necessário definir a relação entre os objetos anatômicos citados e as suas respectivas medidas. Como exemplo, pode-se referir a existência ou não de pontos de calcificação nas artérias coronárias, bem como a quantificação de seus respectivos tamanhos e intensidades.

Em uma recente revisão sistemática sobre o tema, Pons et al. (2016) avaliaram 67 trabalhos que empregaram o uso de PLN em textos médicos. Com base na análise desses artigos foi possível determinar uma sequência de etapas comumente utilizadas para a produção de texto estruturado, a partir de laudos de radiologia que estão escritos em linguagem natural.

**Figura 2: Etapas de PLN usualmente aplicadas à laudos para extração de informações.**



Fonte: adaptado de Pons et al. (2016).

**Figura 3: Fragmento de laudo.**

<p><b>Informações clínicas:</b></p> <p>História de TBC. Dispneia há 2 meses.</p> <p><b>Técnica:</b></p> <p>Exame realizado em aparelho multislice, sem a infusão do meio de contraste, com posteriores reformatações multiplanares, registrados em filmes e CD.</p> <p><b>Interpretação:</b></p> <p>Assimetria dos campos pulmonares, tendo dimensões menores à esquerda. Alterações fibroatelectásicas com bronquiectasias de tração no lobo superior esquerdo, na língula e no segmento superior do lobo inferior homolateral, com espessamento pleural associado.</p>
--

Fonte: Elaborado pelo autor

Conforme apresentado na Figura 2, a primeira etapa, da sequência de etapas identificadas, é denominada segmentação, e consiste na identificação e divisão das seções que podem existir em um laudo. Isso pode contribuir nos passos posteriores do processo, uma vez que cada parte do laudo pode ter características específicas, que devem ser tratadas de maneiras diferentes. A Figura 3 demonstra um trecho de laudo que contém três seções diferentes. É possível que somente a seção que aborda a interpretação do médico seja relevante para a aplicação, então não há motivos para processar o texto por completo.

Uma possível forma de execução dessa tarefa é através de um algoritmo composto por duas etapas. A primeira utiliza um conjunto de regras pré-determinadas, buscando identificar as seções a partir da identificação de títulos e subtítulos, definidos após a análise de um conjunto de laudos. Isso é possível pois, dentro de uma mesma instituição, os textos médicos costumam seguir um padrão estrutural para cada modelo de documento. A segunda etapa compreende a detecção de seções que não possuem um título ou rótulo e é baseada no reconhecimento de padrões léxicos e em análise estatística. Se um número grande de laudos possui cinco seções, caso seja processado um laudo que possui apenas quatro, é provável que a seção que está faltando esteja presente no texto de forma implícita (CHO; TAIRA; KANGARLOO, 2003).

A próxima etapa trata da detecção de limites, cujo resultado é o texto separado em frases ou em palavras. O processo de separação do texto em sentenças é denominado separação de frases (*sentence splitting*) e a separação em palavras é o processo de *tokenização*. A principal indicação de fim de um termo, ou *token*, é o espaço em branco localizado antes e depois do mesmo. Além disso, a palavra pode ser separada também por sinais de pontuação, como ponto, vírgula, ponto e vírgula, ponto de interrogação ou ponto de exclamação. Entretanto, podem ocorrer problemas com termos compostos, como a representação de valores. Como exemplo, pode ser citado o total de leucócitos em um hemograma, representado pelo valor 3.780/uL (AGAH, 2013).

Após a detecção de limites, é executada a etapa de normalização, que pode ser composta pelos processos de remoção de *stopwords*, lematização, *stemming*, correção de erros gramaticais e expansão das abreviações. Todos os passos executados nessa etapa visam eliminar as diferentes formas de escrita de um mesmo termo, reduzindo o número de palavras únicas no texto. *Stopwords* são palavras que não são relevantes para dar sentido à frase, como artigos e preposições. A remoção desse grupo de palavras é feita, usualmente, a partir da deleção de todos os termos que constam em uma lista pré-definida. Existem também estudos que abordam a definição automática dos termos que podem ser removidos (ARMANO; FRANNI; GIULIANI, 2015).

A lematização é a tarefa de determinar a forma base, denominado lema, de uma palavra. Esse método transforma todos os verbos para o tempo infinitivo e converte os substantivos para o singular. Para executar essa parte do processo, Ingason et al. (2008) propuseram um algoritmo que combina aprendizado de máquina com regras baseadas em conhecimento linguístico. Foi adotada essa estratégia híbrida pois, segundo o autor, como um idioma tem inúmeros termos e formas de escrita, é improvável que os dados usados na etapa de treinamento contenham informações suficientes para atender a maior parte dos casos.

Honorato (2008) descreve o processo de *stemming* como a substituição das distintas inflexões de um termo pelo seu respectivo radical. Exemplificando, as palavras “considerar”, “considerado” e “consideração” seriam substituídas, após a execução do método, pelo radical “consider”. Um método amplamente utilizado é o Algoritmo de Porter, em que é verificada a possibilidade de remover os sufixos dos termos. É composto por um conjunto de

aproximadamente 60 sufixos, duas regras de recodificação e uma regra sensível ao contexto, que são aplicadas ao longo de cinco etapas. A primeira envolve o tratamento dos sufixos flexionais (que não modificam a classe gramatical da palavra), as três seguintes avaliam os sufixos derivacionais (que podem alterar a classe gramatical da palavra) e a última é a etapa que emprega a regra de recodificação (WILLETT, 2006).

A etapa seguinte abrange a análise sintática, onde os componentes das frases são classificados de acordo com a estrutura gramatical, relacionamentos ou dependências com outros termos e os tipos de palavras, como, por exemplo, substantivo, verbo ou adjetivo (GÜNGÖR, 2010). Lopes e Alípio (2000) indicam que a complexidade dessa tarefa se dá ao fato de que um mesmo termo pode assumir diferentes papéis dependendo do contexto em que foi utilizado. Os autores propuseram um método misto para resolução desse processo, que consiste na aplicação, de maneira iterativa, de dois tipos de técnicas. A primeira utiliza aprendizado de regras para classificar as palavras. No final do processo, quando restaram apenas as palavras cuja determinação é complexa, são feitas tentativas a partir da execução de um algoritmo de aprendizado baseado em casos, que faz o uso de situações semelhantes para chegar à definição apropriada.

Posteriormente, a etapa de análise semântica é responsável por dar significado às frases e palavras, determinando o seu tipo semântico, como, por exemplo, definindo se uma frase indica a descrição de um sintoma ou de um procedimento. Para este fim, usualmente é utilizado um dicionário de definições e sinônimos, denominado léxico (*lexicon*, em inglês), como o *Unified Medical Language System* (UMLS<sup>1</sup>) ou RadLex<sup>2</sup>, em conjunto com ontologias, que descrevem as relações entre os termos. Abacha e Zweigenbaum (2011) indicam que melhores resultados são obtidos nessa etapa se forem utilizadas abordagens híbridas, que combinam as vantagens dos métodos baseados em regras com as vantagens dos métodos estatísticos.

O último passo antes da utilização efetiva dos dados obtidos compreende o processo de detecção de negação dos conceitos identificados nos passos anteriores. Esta etapa final é especialmente importante, pois, segundo Agah (2013), aproximadamente metade das observações presentes no texto descrevem a ausência de determinada característica no paciente. Desta forma, é essencial que as frases sejam classificadas corretamente, entre afirmações e negações, para que não ocorram falsos-positivos. Para executar essa tarefa é necessário identificar termos específicos utilizados para representar a negação de alguma informação e o contexto em que essa palavra está inserida. Na maior parte dos casos, poucos termos são suficientes para determinar se a característica está ou não presente, porém existem ocasiões em que isso não basta, principalmente se a distância em que o termo que indica a negação é grande em relação à característica correspondente. Os métodos mais simples utilizam regras pré-determinadas e analisadores de padrões léxicos, mas ao longo do tempo estão sendo exploradas abordagens estatísticas e de aprendizado de máquina (OU; PATRICK, 2015).

Além de dados representados na forma de texto, descritos nesta seção, outro grupo de informação amplamente utilizado na área médica é composto por imagens clínicas, que representam visualmente aspectos das condições do paciente. A seção abaixo descreve abordagens que podem ser utilizadas na obtenção de elementos de diagnóstico a partir do processamento dessas imagens.

---

<sup>1</sup> UMLS – <https://www.nlm.nih.gov/research/umls/>

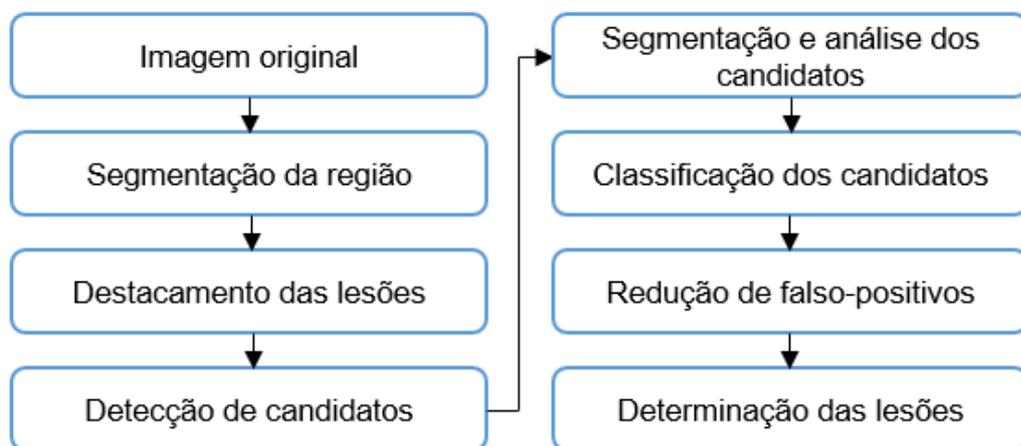
<sup>2</sup> RadLex – <https://www.rsna.org/RadLex.aspx>

## 2.2 Processamento de imagens

Nas últimas três décadas foram realizados rápidos avanços nas tecnologias relacionadas à geração de imagens médicas, com a obtenção de figuras tridimensionais e de alta resolução, possibilitando evoluções poderosas nos processos de diagnóstico e intervenções clínicas. Em conjunto com as melhorias nas imagens, também estão em evolução as aplicações de suporte que fazem uso desses dados. Para essa finalidade, são empregadas diferentes técnicas de processamento digital de imagens (PDI) em operações de redução de ruídos, segmentação de estruturas, extração de características, classificação e visualização (DOUGHERTY, 2009; HANDELS et al., 2013).

Suzuki (2012) apresentou um esquema genérico de passos empregados em sistemas de detecção (CADE). Segundo o autor, esse esquema consiste em quatro etapas que sempre são executadas (segmentação da região de interesse, detecção de elementos candidatos que estão dentro da região segmentada, segmentação e análise dos elementos candidatos e, por fim, classificação dos elementos candidatos) e duas opcionais (destacamento das lesões e redução dos resultados falso-positivos), conforme ilustrado na Figura 4. Todos os passos citados envolvem o emprego ou uso dos resultados da aplicação de técnicas de PDI.

**Figura 4: Esquema genérico de sistemas CADe.**

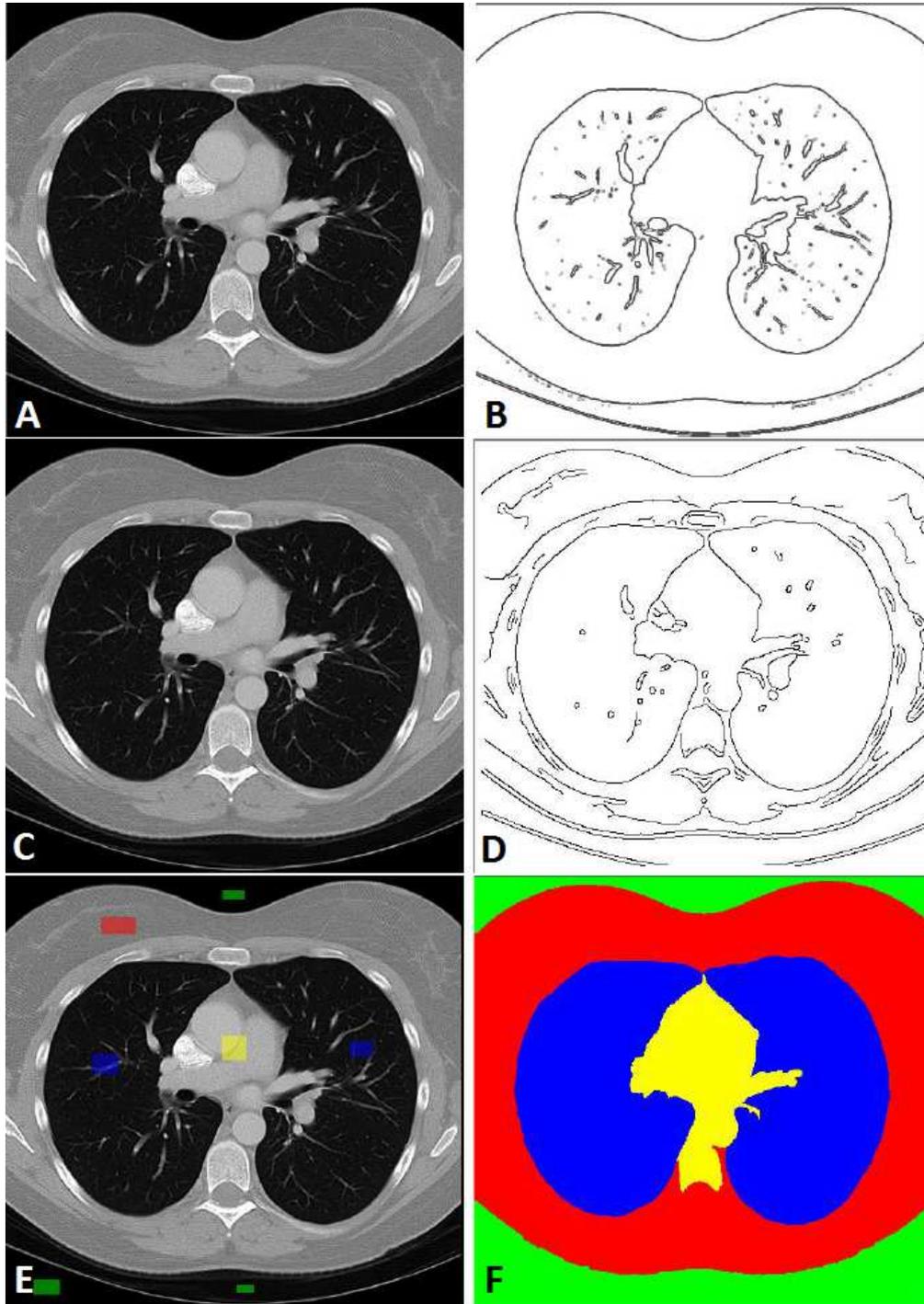


Fonte: adaptado de Suzuki (2012).

Após a obtenção da imagem, representada pela primeira etapa, é executado o processo que delimita a região de interesse a ser avaliada nas operações seguintes. Conforme Gonzalez e Woods (2010), esse processo é denominado segmentação e consiste na divisão da imagem em regiões ou objetos, de modo que seja possível identificar e diferenciar os elementos que se deseja do restante. A operação de segmentação é baseada em duas propriedades: descontinuidade e similaridade. A descontinuidade visa a separar a imagem de acordo com mudanças bruscas dos níveis de cinza, o que caracteriza, por exemplo, a borda de um objeto. Alguns métodos que seguem essa abordagem são o algoritmo de detecção de borda de Sobel (GAO et al., 2010) e o algoritmo de detecção de borda de Canny (CANNY, 1986). Já a similaridade tem como objetivo unir os *pixels* da imagem que apresentam valores e características semelhantes (PEDRINI; SCHWARTZ, 2008). Um exemplo de método que segue essa abordagem é o crescimento de regiões, que separa os objetos contidos em uma imagem, desde que eles tenham características em comum, com baixa tolerância a variações.

As regiões também devem ter a borda bem definida e fechada, de modo que o conteúdo do seu interior esteja separado do restante da imagem (CONCI; AZEVEDO; LETA, 2008).

**Figura 5: Aplicação de técnicas de segmentação em imagem de TC do tórax.**



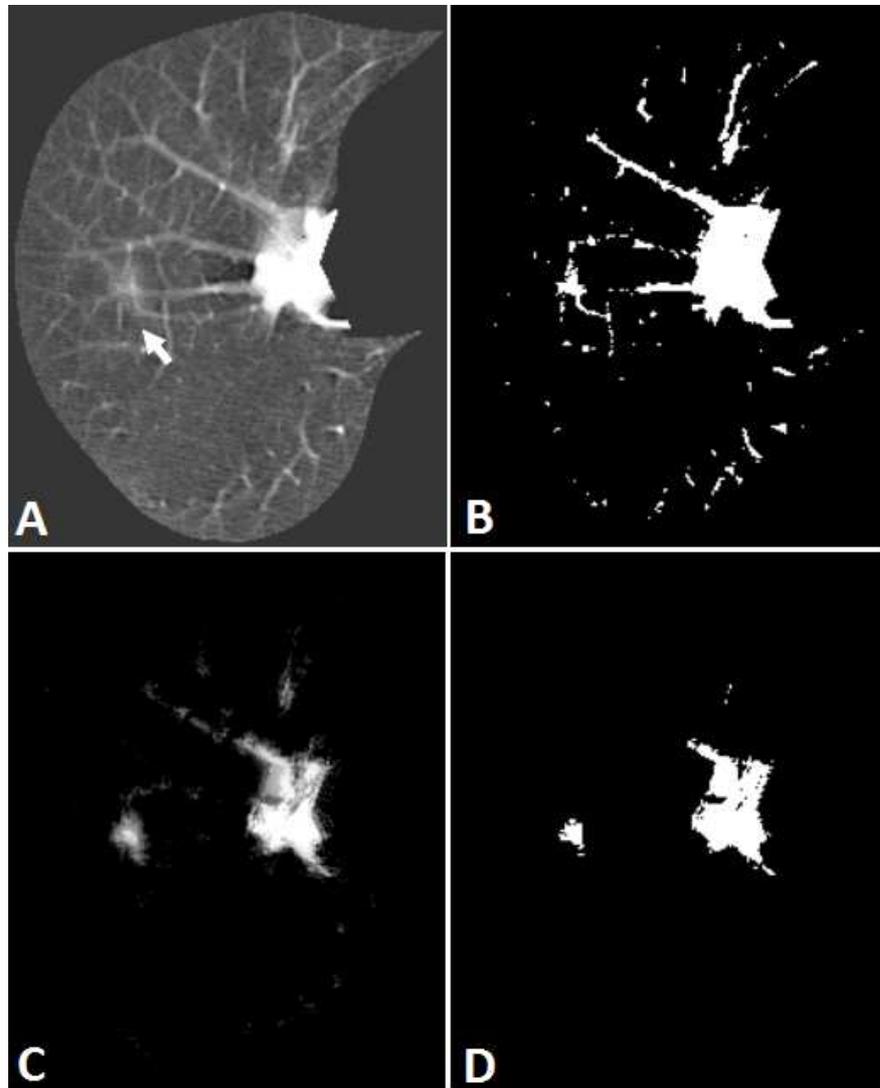
Fonte: Rönnau (2015).

A Figura 5 demonstra os diferentes resultados obtidos a partir da aplicação das técnicas citadas em uma mesma imagem de tomografia computadorizada do tórax. As imagens A e C são as imagens originais, a imagem B é o resultado da aplicação do algoritmo de Sobel, a imagem D é o resultado da aplicação do algoritmo de Canny, a imagem E é a

definição das sementes para o algoritmo de crescimento de regiões e a imagem F é o resultado final da aplicação desse algoritmo.

A etapa seguinte, que é utilizada em parte dos sistemas CADe, é o destacamento das lesões, cuja finalidade é aumentar a acurácia do processo. Isso é obtido com o destacamento dos objetos que possuem características desejáveis, melhorando a sua diferenciação em relação aos eventuais ruídos e, conseqüentemente, aprimorando a identificação. Podem existir diferentes estruturas anatômicas com cor, intensidade e formato semelhante às lesões que se deseja identificar. Por essa razão, a aplicação de uma etapa para diferenciar as lesões dos outros elementos pode melhorar consideravelmente os resultados obtidos pelo sistema. Suzuki (2009) propôs um modelo de filtro supervisionado para destacar nódulos pulmonares em exames de TC, fazendo uso de uma rede neural artificial, que é capaz de distinguir lesões com níveis de opacidade específicos.

**Figura 6: Resultado das etapas de destacamento das lesões e detecção de candidatos.**



Fonte: adaptado de Suzuki (2009).

Posteriormente é feita a identificação de elementos que serão considerados candidatos. Essas estruturas possuem características básicas que as diferenciam das demais estruturas, podendo ser delimitadas a partir de sua tonalidade. Para isso, é definido um valor de limiar e

aplicado à imagem. A limiarização consiste em comparar cada um dos *pixels* que compõem a figura com um valor de limiar (também denominado *threshold*, em inglês) e, dependendo do resultado dessa comparação, determinar um novo valor para o pixel avaliado. A Figura 6 apresenta o resultado das etapas de destacamento das lesões e detecção de candidatos. A imagem A representa a região do pulmão já segmentada, a imagem B é o resultado da etapa de identificação de candidatos somente com a aplicação do limiar. A imagem C representa o resultado da etapa de destacamento das lesões. É possível perceber que alguns elementos já foram descartados. A imagem D é o resultado da identificação de elementos candidatos após o destacamento das lesões.

Para isolar os elementos definidos como candidatos, possibilitando a posterior análise individual, pode ser utilizado um algoritmo de rotulação de objetos. He, Chao e Suzuki (2008) indicam um método para execução dessa tarefa em imagens binárias, ou seja, que possuem somente dois valores possíveis para os *pixels*. Nesse caso, uma tonalidade define o fundo da imagem e a outra define os objetos. O processo consiste em percorrer todos os *pixels*, atribuindo um rótulo para cada um. Todos os pontos que estiverem conectados ao ponto avaliado recebem o mesmo rótulo. Dessa forma, ao final do processo, cada objeto da imagem estará com um rótulo diferente e todos os pontos que constituem cada objeto estarão identificados. Na análise dos objetos candidatos as características morfológicas, texturas e ou informações de cada elemento são extraídas. Essas informações são empregadas então nas etapas seguintes, em que os candidatos são classificados. A determinação se um dado objeto representa efetivamente uma lesão é feita, usualmente, com o uso de classificadores. Essa etapa tem como objetivo remover o máximo de objetos que não são lesões e evitar a remoção de objetos que são lesões. Geralmente, são utilizadas regras pré-determinadas ou técnicas de aprendizado de máquina, que fazem uso do conjunto de características que foi extraído no passo anterior. A redução de falso-positivos também é baseada, usualmente, em técnicas de aprendizado de máquina (SUZUKI, 2012).

**Figura 7: Aplicação de filtro Gaussiano com máscara 3x3 em imagem de TC do tórax.**



Fonte: Rönnau (2015).

Segundo Verma e Khanna (2016), outra importante etapa que é usualmente executada em sistemas de apoio é o pré-processamento, que abrange diferentes operações de preparação das imagens. Essa etapa tem como objetivo melhorar a imagem e facilitar a execução das etapas posteriores. São aplicados, comumente, algoritmos para realçar as estruturas de

interesse, a partir da remoção ou atenuação de ruídos, ajuste de brilho e contraste, entre outros (BANKMAN, 2000). A remoção de ruídos pode ser obtida através da aplicação de filtros, de modo que cada *pixel* que forma a imagem é avaliado, em conjunto com seus vizinhos, podendo ter seu valor modificado. Um exemplo de filtro amplamente utilizado é o Gaussiano (CONCI; AZEVEDO; LETA, 2008), cujo resultado é ilustrado na Figura 7.

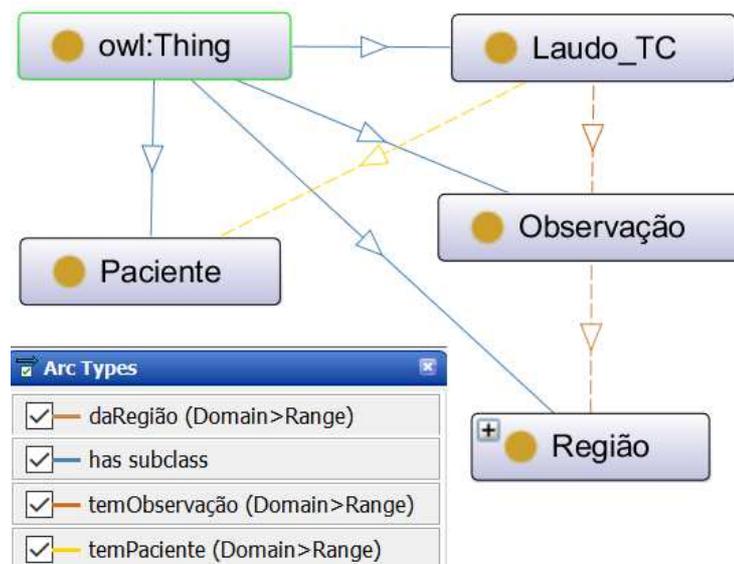
Além das técnicas citadas, existem diversas outras que podem ser exploradas e aplicadas, de acordo com a finalidade em que são empregadas. Usualmente, o conjunto de métodos de PDI que são utilizadas depende do objetivo da aplicação. Estas não serão descritas no presente trabalho pois não estão relacionadas diretamente ao seu escopo, mas podem ser encontradas mais informações a respeito na literatura da área, como em Gonzalez e Woods (2010), Conci, Azevedo e Leta (2008) e Pedrini e Schwartz (2008).

Para que seja possível utilizar as características extraídas a partir do processamento de diferentes fontes de dados clínicos, como textos e imagens, em aplicações computacionais, é necessário que os dados contenham detalhes suficientes e que estejam armazenados de maneira estruturada. A seção a seguir apresenta o uso de ontologias para essa finalidade.

## 2.3 Ontologias

Uma das formas para representação de conhecimento é o uso de ontologias, de modo que, segundo Russell e Norvig (2013), as informações estejam organizadas de maneira hierárquica e em categorias. Essa estrutura é determinada por especialistas, que indicam as regras existentes entre os termos e suas respectivas relações, de acordo com o domínio de conhecimento que se deseja modelar. Os conceitos, propriedades, relações, funções, restrições e axiomas devem ser definidos por especialistas, dentro de um contexto reconhecido como válido para uma comunidade de usuários (ALMEIDA; BAX, 2003).

**Figura 8: Exemplo de ontologia utilizada para representar o conteúdo de laudos médicos.**



Fonte: Elaborado pelo autor.

Um exemplo de ontologia utilizada na área médica é apresentado na Figura 8, em que classes, interações e atributos referentes ao conteúdo de laudos médicos são definidos. Esse

formato permite a inserção das informações que constam em textos clínicos no modelo, de maneira estruturada, mantendo as relações entre os conceitos existentes.

A classe owl:Thing é o item que generaliza os demais, agrupando todos os conceitos filhos. A classe “Laudo\_TC” tem os atributos “codigoExame”, “contraste”, “corte” e “dataExame”. Além disso, possui uma relação denominada “temPaciente” com a classe “Paciente” e pode ter múltiplos relacionamentos denominados “temObservação” com a classe “Observação”. A classe “Paciente”, por sua vez, é composta pelos atributos “codigoPaciente”, “dataNascimento”, “nome” e “sexo”. Já a classe “Observação” tem o atributo “característica” e o relacionamento “daRegião”, que pode fazer referência a qualquer subclasse da classe “Região”. A classe “Região” pode ser a representação de uma ligação com outra ontologia que descreve, por exemplo, a estrutura anatômica do corpo humano, como a disponibilizada pelo *National Cancer Institute* no BioPortal<sup>3</sup>, denominada NCIT<sup>4</sup> (*National Cancer Institute Thesaurus*).

A organização da ontologia utilizada como exemplo segue uma estrutura hierárquica (taxonomia), cujas relações representam o tipo de iteração entre as classes. As instâncias (que não foram explicitadas na Figura 8) de cada uma das classes representam os dados. A partir desse modelo é permitido realizar a extração de dados com a execução de consultas diretas ou baseadas na estrutura de conceitos, que fazem o uso das relações explícitas e implícitas. Esse tipo de consulta pode ser descrito utilizando a linguagem SPARQL (*SPARQL Protocol and RDF Query Language*). Em uma consulta SPARQL é possível selecionar os dados a serem retornados, definir filtros e condições, bem como escolher a fonte de dados a ser comparada com o padrão informado na busca (PÉREZ; ARENAS; GUTIERREZ, 2006).

A forma como os dados são estruturados em uma ontologia, de maneira hierárquica, é semelhante ao empregado por determinados padrões atualmente utilizados na área médica. A próxima seção descreve alguns desses padrões, possibilitando a identificação de equivalências em relação às ontologias.

## 2.4 Padrões de representação de dados clínicos

O tipo de representação escolhido para armazenar as informações obtidas nos processos de extração de características deve atender especificações e padrões atualmente estabelecidos, bem como ser compatível com outros sistemas médicos, garantindo assim a interoperabilidade entre diferentes aplicações. O EHR agrupa diferentes tipos de informações, como descrições de consultas, procedimentos médicos, medicamentos utilizados e resultados de exames, podendo ser empregado em aplicações de suporte ao paciente, pesquisa e ensino (ARAUJO; PIRES; BANDIERA-PAIVA, 2014).

Existem diversos tipos de padrões cuja finalidade é prover uma maneira organizada de representar e manipular dados clínicos. Destes, o Ministério da Saúde do Brasil determinou, através da Portaria nº 2.073 (BRASIL, 2011), quais devem ser aplicados em sistemas de saúde utilizados no país, de acordo com as seguintes indicações:

- a) openEHR deve ser utilizado como modelo de referência para a definição do EHR;

---

<sup>3</sup> BioPortal – <http://bioportal.bioontology.org/>

<sup>4</sup> NCIT – <http://bioportal.bioontology.org/ontologies/NCIT/>

- b) HL7 deve ser utilizado para estabelecer a interoperabilidade entre sistemas (integração dos resultados e solicitações de exames);
- c) SNOMED-CT deve ser empregado na codificação de termos clínicos e mapeamento das terminologias;
- d) HL7 CDA deve ser utilizado como arquitetura de documentos clínicos;
- e) DICOM deve ser utilizado para representar informações relativas a exames de imagem.

As seções seguintes apresentam detalhes sobre os padrões citados, expandindo aqueles que não estão especificados na portaria descrita, mas que são empregados em sistemas utilizados por outros países.

#### 2.4.1 OpenEHR

O padrão openEHR tem como objetivo garantir a interoperabilidade universal entre dados de saúde, com foco em EHR e sistemas relacionados. Para alcançar esse objetivo foi definida uma arquitetura de software orientada a serviço, que é formada por diferentes níveis. O primeiro nível contém o modelo de referência, que define a estrutura lógica dos dados. Qualquer informação no EHR deve estar de acordo com o que é especificado nesse modelo. O nível seguinte consiste em um grupo de arquétipos, que são estruturas de dados que descrevem conceitos clínicos. A biblioteca internacional de arquétipos (CKM<sup>5</sup>) contém aproximadamente 500 elementos, eliminando assim a necessidade de recriar estruturas que já foram definidas e validadas. Só devem ser modelados os arquétipos não existentes no CKM. Posteriormente, os arquétipos são agregados de acordo com o contexto de uso, formando assim os *templates*. Todos os sistemas que utilizam o padrão openEHR são desenvolvidos com base nos *templates*. A Fundação openEHR (responsável pelo padrão) disponibiliza ferramentas para desenvolvimento de arquétipos e *templates*. O último nível é formado pelos artefatos gerados a partir dos *templates*, que são utilizados pelos desenvolvedores e usuários em cada aplicação (OPENEHR, 2017a).

Segundo Bacelar e Correia (2015), um arquétipo deve representar um conceito clínico de forma estruturada e da maneira mais detalhada possível, de modo que esteja preparado para ser usado em diferentes aplicações. Um arquétipo é definido a partir da linguagem ADL (*Archetype Definition Language*), contendo somente tipos de dados que estão descritos no modelo de referência, como texto, quantidade ou data. O significado do arquétipo não é alterado quando os termos que o compõem e o descrevem são traduzidos, possibilitando assim a troca de informações entre sistemas que estão em idiomas distintos sem que ocorram perdas de conteúdo. O mesmo ocorre em relação às múltiplas terminologias (como SNOMED-CT) que podem ser associadas a cada estrutura. A Figura 9 ilustra a arquitetura do openEHR, indicando as interações entre os conceitos descritos.

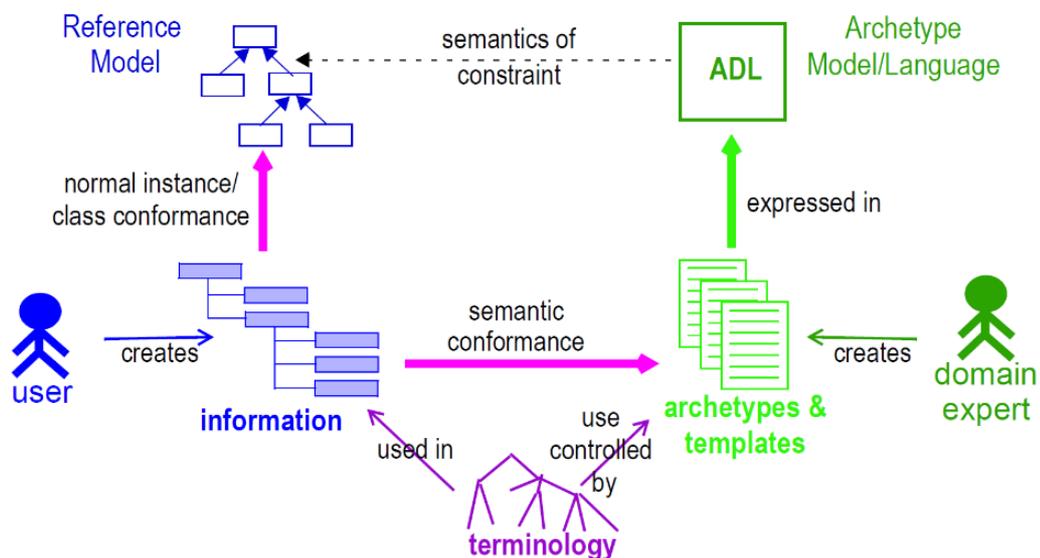
Os arquétipos utilizados para representar dados clínicos podem ser divididos em quatro grandes grupos, que são utilizados de acordo com a finalidade do *template* em que são empregados. Os tipos existentes são:

---

<sup>5</sup> CKM – <http://www.openehr.org/ckm/>

- a) Observação (*observation*): agrupa as descrições feitas pelo paciente, relacionadas aos sintomas, eventos e preocupações, além dos resultados de exames;
- b) Avaliação (*evaluation*): agrupa as interpretações, opiniões e diagnósticos dos profissionais obtidas a partir da análise das observações;
- c) Instrução (*instruction*): agrupa as prescrições, recomendações e solicitações de exames referentes às observações e avaliações feitas;
- d) Ação (*action*): agrupa as atividades executadas a partir das instruções dadas, como por exemplo, a realização de procedimentos cirúrgicos ou administração de medicamentos.

Figura 9: Arquitetura do openEHR.



Fonte: openEHR (2016).

O *template* é a especificação que define uma árvore de um ou mais arquétipos, sendo estes conjuntos de uma ou mais instâncias de tipos de dados definidos no modelo de referência. Um *template* é usado em tempo de execução para criar estruturas e para validar a entrada de dados, garantindo que todas as informações inseridas estejam de acordo com as restrições definidas nos arquétipos utilizados, em relação aos tipos de dados e terminologias empregadas (OPENEHR, 2016).

A linguagem de consulta desenvolvida especificamente para ser utilizada em arquétipos é denominada AQL (*Archetype Query Language*), cuja sintaxe é independente das aplicações, linguagens de programação utilizadas e modelos de armazenamento. O comando de pesquisa pode ser formado por cinco cláusulas: “SELECT”, “FROM”, “WHERE”, “ORDER BY” e “TIMEWINDOW”. A primeira é a especificação dos dados a serem recuperados, a segunda indica a origem das informações, a terceira as condições para filtrar o conteúdo retornado, a quarta é a definição da ordem em que os resultados serão apresentados e a última a delimitação do escopo de tempo a ser avaliado (OPENEHR, 2017b). Um exemplo de consulta que retorna valores de pressão sanguínea é apresentado na Figura 10.

Figura 10: Exemplo de consulta utilizando AQL.



Fonte: openEHR (2017b).

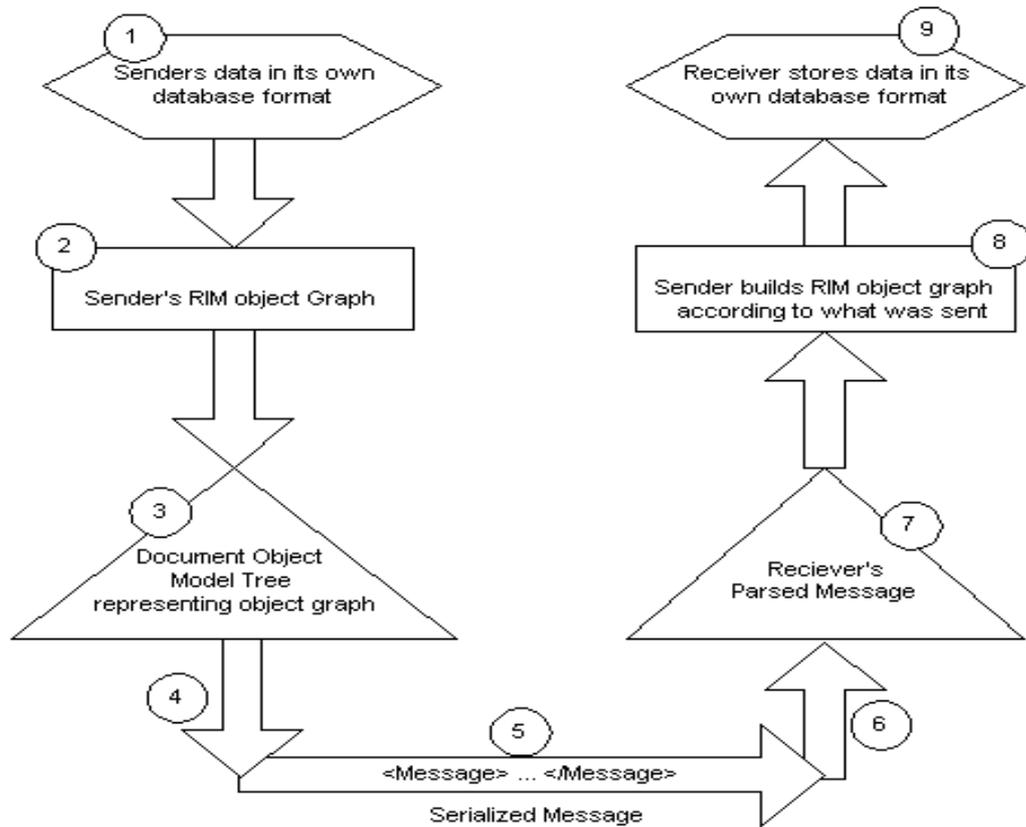
#### 2.4.2 HL7

HL7 (*Health Level 7*) é um conjunto de padrões que especificam uma estrutura para o intercâmbio, integração, compartilhamento e recuperação de informações clínicas. Dentre as especificações, está a determinação de como os registros são empacotados e comunicados de uma parte para outra, definindo a linguagem, a estrutura e os tipos de dados necessários para a integração entre diferentes sistemas (HL7, 2017a). Dentre os padrões existentes, destacam-se o HL7 v3, o HL7 CDA e o HL7 FHIR.

O HL7 v3 (*Health Level Seven Version 3*) é um conjunto de especificações utilizadas com o objetivo de possibilitar o intercâmbio de informações clínicas a partir do uso de mensagens e documentos expressos em XML. São definidas diferentes características, como o formato das mensagens, tipos de dados (textos, códigos, identificadores, nomes, endereços, quantidades) e terminologias necessárias para o desenvolvimento das aplicações (HL7, 2017b). O *Reference Information Model* (RIM) é a base das estruturas utilizadas nas mensagens HL7, descrevendo as classes, seus respectivos atributos e relações, bem como os tipos de dados que cada atributo pode conter e os vocabulários aceitos. O fluxo do envio e recebimento de uma mensagem HL7 é ilustrado na Figura 11.

Os itens 1 e 9 demonstram que os dados podem ser armazenados nas bases do remetente e destinatário em um formato diferente do estabelecido pelo padrão HL7. Os itens 2, 3, 7 e 8 ilustram o uso do RIM para montar e reconstruir a mensagem, que foi convertida para o padrão XML nos itens 4 e 6. A transmissão, representada no item 5, pode ocorrer de diferentes maneiras, como, por exemplo, por e-mail. A *Implementation Technology Specification* (ITS) define como os objetos RIM são representados nas mensagens. O modelo recomendado no padrão HL7 é o XML, mas estão em desenvolvimento outros formatos, como o UML (*Unified Modeling Language*). A Figura 12 apresenta um fragmento de mensagem no padrão HL7 utilizando XML ITS. Nesse fragmento, consta o profissional designado para executar o exame, contendo sua identificação, nome e sufixo.

**Figura 11: Fluxo do envio e recebimento de uma mensagem HL7 v3.**



Fonte: HL7 (2017b).

**Figura 12: Fragmento de uma mensagem padrão HL7 utilizando XML ITS.**

```

<participant>
  <!-- this is the signatory -->
  <type_cd code="AUT"/>
  <!-- time of signature -->
  <time value="2002-08-16T09:30:00"/>
  <mode_cd code="WRITTEN"/>
  <!-- signature is on file -->
  <signature_cd code="S"/>
  <assignedEntity>
    <id root="2.16.840.1.113883.1122"
      extension="444-444-4444"/>
    <assignee_Person>
      <nm use="L" xsi:type="dt:PN">
        <dt:family>Hippocrates</dt:family>
        <dt:given>Harold</dt:given>
        <dt:given>H</dt:given>
        <dt:suffix qualifier="AC">MD</dt:suffix>
      </nm>
    </assignee_Person>
  </assignedEntity>
</participant>

```

Fonte: Adaptado de HL7 (2017b).

O HL7 *Clinical Document Architecture* (CDA) é um padrão de representação de documentos clínicos, como resumos de alta ou laudos de exames, que tem como propósito especificar a sua estrutura e semântica para operações de troca de informações entre instituições de saúde, prestadores de serviços e pacientes. Um documento CDA é codificado no padrão XML, podendo conter textos, imagens, sons e outros conteúdos. É baseado também no RIM e utiliza os tipos de dados do HL7 v3. É possível transmitir documentos nesse padrão a partir de mensagens HL7 (HL7, 2017c).

Um documento padrão CDA é composto por diversos elementos que indicam o tipo do conteúdo que está entre suas indicações de abertura (<exemplo>) e de fechamento (</exemplo>). Todo o documento é envolvido pelo elemento “<ClinicalDocument>”, sendo dividido posteriormente em cabeçalho e corpo. O cabeçalho contém a identificação e classificação do documento, fornecendo informações sobre a autenticação, o encontro, o paciente e os profissionais envolvidos. Já o corpo é composto pelos dados clínicos, podendo estar organizado de maneira estruturada ou não estruturada. Quando se opta por representar as informações de maneira estruturada é utilizado o elemento “<structuredBody>”. O conteúdo do corpo pode ser composto por inúmeras seções (<section>), de acordo com o que se deseja representar. Em uma seção pode existir um bloco narrativo (<text>), que contém os dados legíveis para as pessoas, e nenhuma ou mais entradas CDA, que contém, usualmente, os mesmos dados existentes no bloco narrativo, mas codificados de forma que seja possível processar a partir de um computador. A Figura 13 apresenta um exemplo da estrutura que um documento construído de acordo com o padrão CDA possui.

**Figura 13: Exemplo da estrutura de um documento padrão CDA.**

```
<ClinicalDocument>
  <!-- ... CDA Header ... -->
  <structuredBody>
    <!-- ... CDA Body ... -->
    <section>
      <text>...</text>
      <observation>...</observation>
      <substanceAdministration>
        <supply>...</supply>
      </substanceAdministration>
      <observation>
        <externalObservation>...
      </externalObservation>
      </observation>
    </section>
    <section>
      <section>...</section>
    </section>
  </structuredBody>
</ClinicalDocument>
```

Fonte: Adaptado de HL7 (2017c).

Podem ser feitas referências a objetos externos, ou seja, que não estão no documento CDA, utilizando o elemento “<externalObservation>”. Isso pode ser usado para relacionar as informações descritas no documento com, por exemplo, uma imagem médica.

É possível definir uma série de modelos (*templates*) para documentos CDA, determinando assim padrões de formato e restrições de conteúdo. Dentro do grupo de modelos, pode-se dividir em dois subgrupos: modelos no nível de seção e modelos no nível de

entrada. O primeiro tipo limita as seções que podem existir dentro de um documento. Já o segundo restringe as entradas que podem existir em determinada seção. A representação em XML de arquivos CDA segue o que foi definido na HL7 v3 XML ITS.

O padrão HL7 *Fast Healthcare Interoperable Resources* (FHIR) foi desenvolvido com foco em implementabilidade, com base nos padrões HL7 v2, v3 e CDA. É formado principalmente por dois tipos de componentes: os *resources* (recursos) e as APIs<sup>6</sup>. Os *resources* são modelos de informação que definem os elementos que representam os dados, as regras, as restrições e os relacionamentos. Como exemplos, podem ser citados modelos utilizados para descrever um paciente, uma observação ou um procedimento. A lista dos 116 tipos de *resources* definidos no padrão, bem como as suas respectivas especificações, se encontra no *Resource Index*<sup>7</sup>. Já as APIs são interfaces bem definidas para interoperação entre duas aplicações. As quatro principais maneiras de transferir informações que estão no padrão HL7 FHIR são: interface REST, transferência de documentos (cujo agrupamento dos elementos é feito com o *resource Composition*), envio e recebimento de mensagens (cujo agrupamento dos elementos é feito com o *resource MessageHeader*) e, por fim, exposição e invocação de serviços. Por se tratar de uma arquitetura modular, os componentes escolhidos e empregados podem variar de acordo com a finalidade do sistema desenvolvido (HL7, 2017d).

Cada *resource* é formado por um conjunto de características, que pode variar de acordo com o tipo de informação que será descrito. A Figura 14 apresenta um exemplo de representação de um paciente com o uso do *resource "Patient"*, em JSON. É possível representar as instâncias desses elementos utilizando diferentes formatos, como XML, JSON<sup>8</sup> e RDF<sup>9</sup>.

Em comum, todos recursos possuem:

- a) Uma URL (*Uniform Resource Locator*) que o identifica;
- b) Um grupo de dados comuns, denominado *common metadata* (linhas 4 a 7 da Figura 14);
- c) Um resumo do recurso descrito na forma de texto não estruturado e legível para seres humanos (linhas 8 a 11 da Figura 14);
- d) Uma estrutura que pode descrever informações adicionais que não fazem parte da definição básica do recurso (linhas 12 a 15 da Figura 14).
- e) Um conjunto de elementos padrões que armazenam os dados representados pelo recurso (linhas 16 a 29 da Figura 14);

Diferentes terminologias podem ser utilizadas nos dados codificados, como as especificadas no HL7 v3, SNOMED-CT (que é descrita na seção a seguir) ou mesmo baseadas em diretórios locais. Ao utilizar a API RESTful, podem ser executadas buscas nas instâncias inseridas a partir da operação GET. Outra maneira de recuperar informações é diretamente na base de dados utilizada, via SQL (*Structured Query Language*).

---

<sup>6</sup> API – *Application program interface*.

<sup>7</sup> *Resource Index* – <http://hl7.org/fhir/resourcelist.html>

<sup>8</sup> JSON – *JavaScript Object Notation*.

<sup>9</sup> RDF – *Resource Description Framework*.

Figura 14: Exemplo de representação de um paciente em HL7 FHIR usando JSON.

```

1  {
2    "resourceType": "Patient",
3    "id" : "23434",
4    "meta" : {
5      "versionId" : "12",
6      "lastUpdated" : "2014-08-18T15:43:30Z"
7    }
8    "text": {
9      "status": "generated",
10     "div": "<!-- removido para reduzir o tamanho do exemplo -->"
11   },
12   "extension": [
13     { "url": "http://example.org/consent#trials",
14       "valueCode": "renal" }
15   ],
16   "identifier": [
17     { "use": "usual",
18       "label": "MRN",
19       "system": "http://www.goodhealth.org/identifiers/mrn",
20       "value": "123456" }
21   ],
22   "name": [
23     { "family": ["Levin"],
24       "given": ["Henry"],
25       "suffix": ["The 7th"] }
26   ],
27   "gender": {"text": "Male" },
28   "birthDate": "1932-09-24",
29   "active": true
30 }

```

Fonte: Adaptado de HL7 (2017d).

### 2.4.3 SNOMED-CT

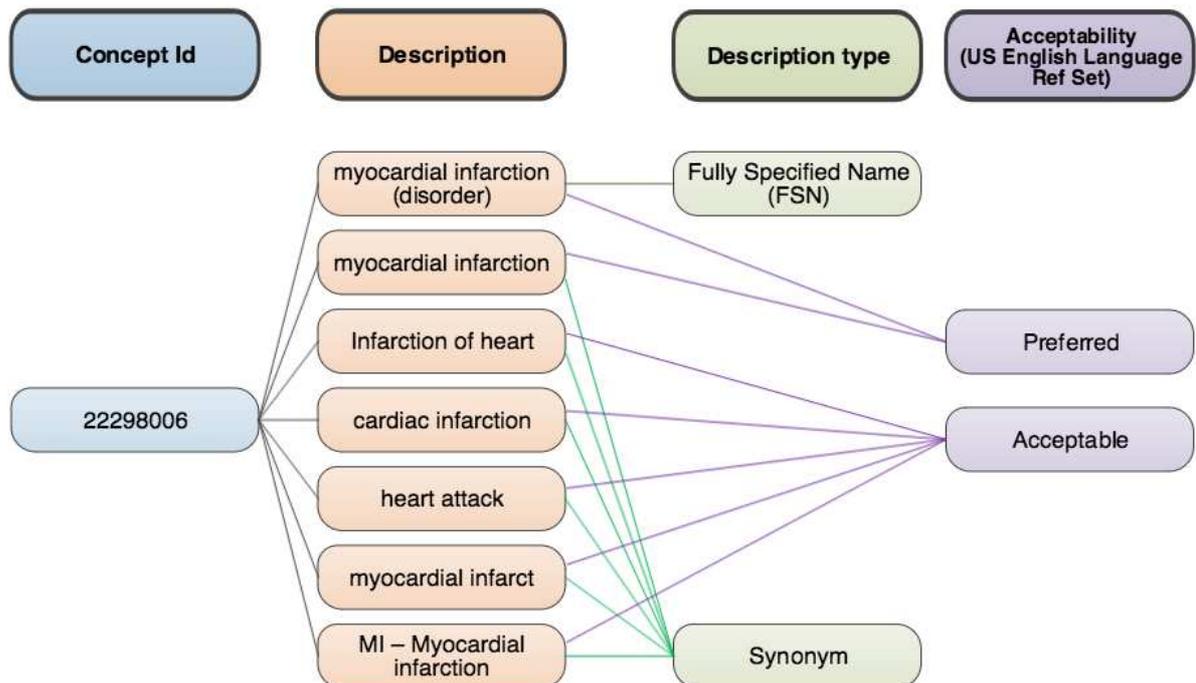
O SNOMED CT é uma terminologia clínica, formada por um conjunto de conceitos únicos que são organizados de maneira hierárquica, cobrindo diversos tópicos relacionados à saúde. É possível empregar esse padrão, por exemplo, na descrição do histórico médico de um paciente ou no detalhamento de um procedimento ortopédico. Dessa forma, as informações são representadas de maneira padronizada, possibilitando o uso de sistemas computacionais para interpretá-las. Além disso, outra vantagem é possibilitar a interoperabilidade entre instituições diferentes, pois as informações estarão sendo representadas de maneira padronizada. A versão liberada em julho de 2016 contém 321901 conceitos ativos (SNOMED, 2017a).

São utilizados três tipos de componentes: conceitos, descritores e relacionamentos. O primeiro tipo representa as informações clínicas através de um identificador numérico único. A organização hierárquica é feita de acordo com o nível de especificação de cada termo. Os descritores são responsáveis pela ligação entre os conceitos e seus respectivos significados, expressados de maneira apropriada para o entendimento de seres humanos. Um mesmo conceito pode ter mais de um descritor, de modo que cada descritor é considerado um

sinônimo. Os descritores também possuem um identificador numérico único. A Figura 15 ilustra um exemplo de diferentes descritores que estão relacionados a um mesmo conceito, bem como o identificador do conceito e indicação de descritores preferenciais.

Os relacionamentos ligam conceitos a outros cujo significado está relacionado de alguma maneira, resultando assim em definições formais e indicando outras propriedades de um termo. Os sistemas que armazenam as informações clínicas passaram a utilizar os identificadores únicos dos conceitos existentes na terminologia SNOMED-CT no lugar de palavras. É possível visualizar toda a estrutura de conceitos através do *SNOMED CT Browser*<sup>10</sup> (SNOMED, 2017b).

Figura 15: Exemplo de diferentes descritores para um mesmo conceito.



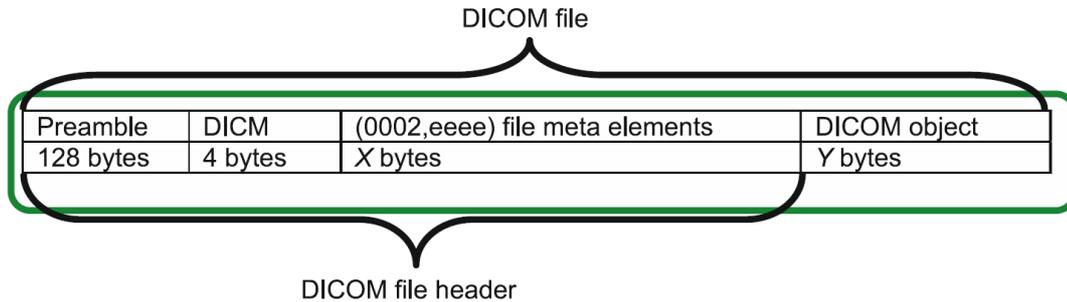
Fonte: SNOMED (2017b).

#### 2.4.4 DICOM

O padrão DICOM (*Digital Imaging and Communications in Medicine*) define características de geração, armazenamento, comunicação e manipulação de imagens médicas e dados relacionados. São especificadas características referentes a rede de comunicação entre os equipamentos, a sintaxe e a semântica dos comandos utilizados no protocolo, a estrutura de dispositivos e os diretórios de armazenamento, além do formato dos arquivos que contêm as imagens e informações dos exames (NEMA, 2017). O formato de arquivos desse padrão possui a estrutura apresentada na Figura 16, sendo composto por um *header*, que é formado por três partes, e um *data set*.

<sup>10</sup> SNOMED CT Browser – <http://browser.ihtsdotools.org>

Figura 16: Estrutura geral de um arquivo no padrão DICOM.

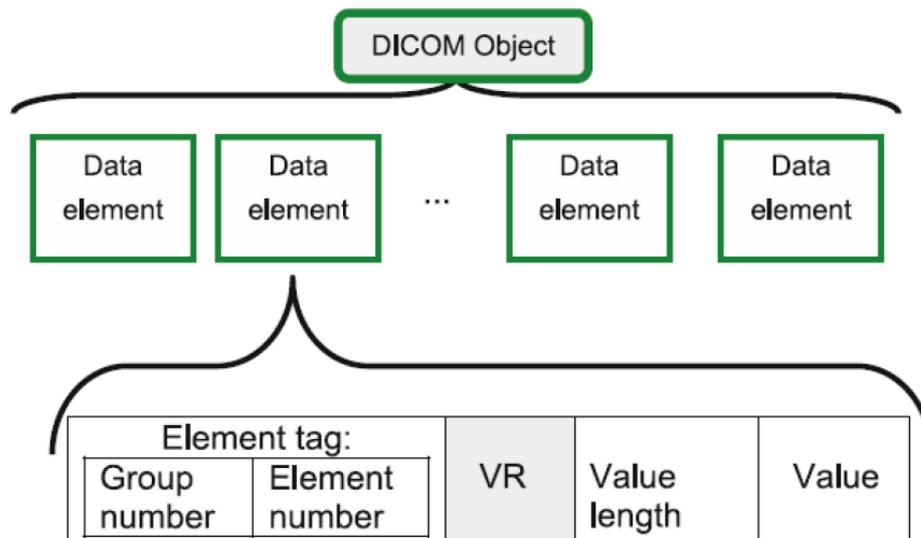


Fonte: Pianykh (2008).

A primeira parte do *header* é uma área de 128 bytes que pode ser usada de maneira livre, denominada *preamble*. Justamente por não ter o conteúdo definido, essa área é, usualmente, ignorada pelas aplicações. A segunda parte contém o prefixo, cuja finalidade é indicar aos programas que o arquivo está no padrão. O conteúdo é sempre a literal “DICM”. Tanto o *preamble* como o prefixo não são codificados, permitindo assim que qualquer aplicação consiga ler e entender o conteúdo (PIANYKH, 2008). A última parte do *header* é o grupo de atributos, ou *file meta elements*, que contém informações do arquivo. Exemplos de dados existentes nessa área são: versão do padrão, identificadores de instâncias, tipo de codificação, origem, destino, entre outros (NEMA, 2017).

Após o cabeçalho estão localizados os dados do objeto DICOM, na forma de *data elements*. A Figura 17 ilustra o formato desse conjunto. Os *meta elements* e os *data elements* possuem, no início da estrutura, uma *tag*, formada pelo número do grupo a qual o elemento pertence e o número de identificação do elemento descrito. O grupo dos *meta elements* é “0002”, conforme pode ser observado na Figura 16. O grupo que define o início dos *data elements* é “0008”, podendo ser diferente para os elementos seguintes. De acordo com Pianykh (2008), essa diferença possibilita diferenciar os dados que fazem parte do *header* e os dados que fazem parte do objeto.

Figura 17: Conjunto de *data elements* em um arquivo no padrão DICOM.



Fonte: Pianykh (2008).

A outra parte da estrutura é dividida em mais três informações: VR, *value length* e *value*. O primeiro, VR (*value representation*), é composto por dois caracteres e descreve o tipo de dado e o formato do valor do elemento. Exemplos de VR são: “AS” quando for desejado representar uma idade; “DA” para representar uma data; “DS” para valores decimais, “IS” para valores inteiros. O *value length* indica o número de bytes utilizados para armazenar o valor, podendo ser formado por 16 ou 32 bits, de acordo com o VR especificado. Por fim, o *value* é o dado, ou conjunto de dados, cujo identificador é o *element number* (ver Figura 17), o tamanho é indicado pelo *value length* e o tipo pelo VR (NEMA, 2017).

Conforme Pinykh (2008), existem mais de 2000 atributos diferentes, possibilitando a existência de diversos tipos de dados em arquivos no padrão DICOM, como informações do equipamento utilizado na obtenção das imagens, dados referentes ao paciente, aos profissionais envolvidos e do exame. A Tabela 2 contém exemplos de elementos que contêm propriedades das imagens.

**Tabela 2: Exemplos de propriedades referentes às imagens em arquivos no padrão DICOM.**

<b>(grupo,identificador)</b>	<b>Elemento</b>	<b>VR</b>	<b>Descrição</b>
(0018,0050)	<i>Slice Thickness</i>	DS	Espessura de cada imagem da série.
(0018,0088)	<i>Spacing Between Slices</i>	DS	Distância física entre cada imagem da série.
(0028,0008)	<i>Number of Frames</i>	IS	Número de imagens que compõem o estudo.
(0028,0010)	<i>Rows</i>	US	Número de linhas da imagem.
(0028,0011)	<i>Columns</i>	US	Número de colunas da imagem.
(0028,0030)	<i>Pixel Spacing</i>	DS	Tamanho físico de cada pixel da imagem.
(7FE0,0010)	<i>Pixel Data</i>	OB ou OW	Valores dos pixels da imagem.

Fonte: Elaborado pelo autor.

## 2.5 Considerações

É possível classificar os padrões apresentados nas seções anteriores em três grupos, de acordo com suas respectivas aplicações: a) terminologias; b) representação de imagens e informações relacionadas; c) representação de dados genéricos e textuais. No grupo de terminologias está o padrão SNOMED-CT. O padrão utilizado para representação e transmissão de imagens médicas é o DICOM. Por fim, no grupo de representação de dados diversos estão os padrões openEHR, HL7 v3, HL7 CDA e HL7 FHIR.

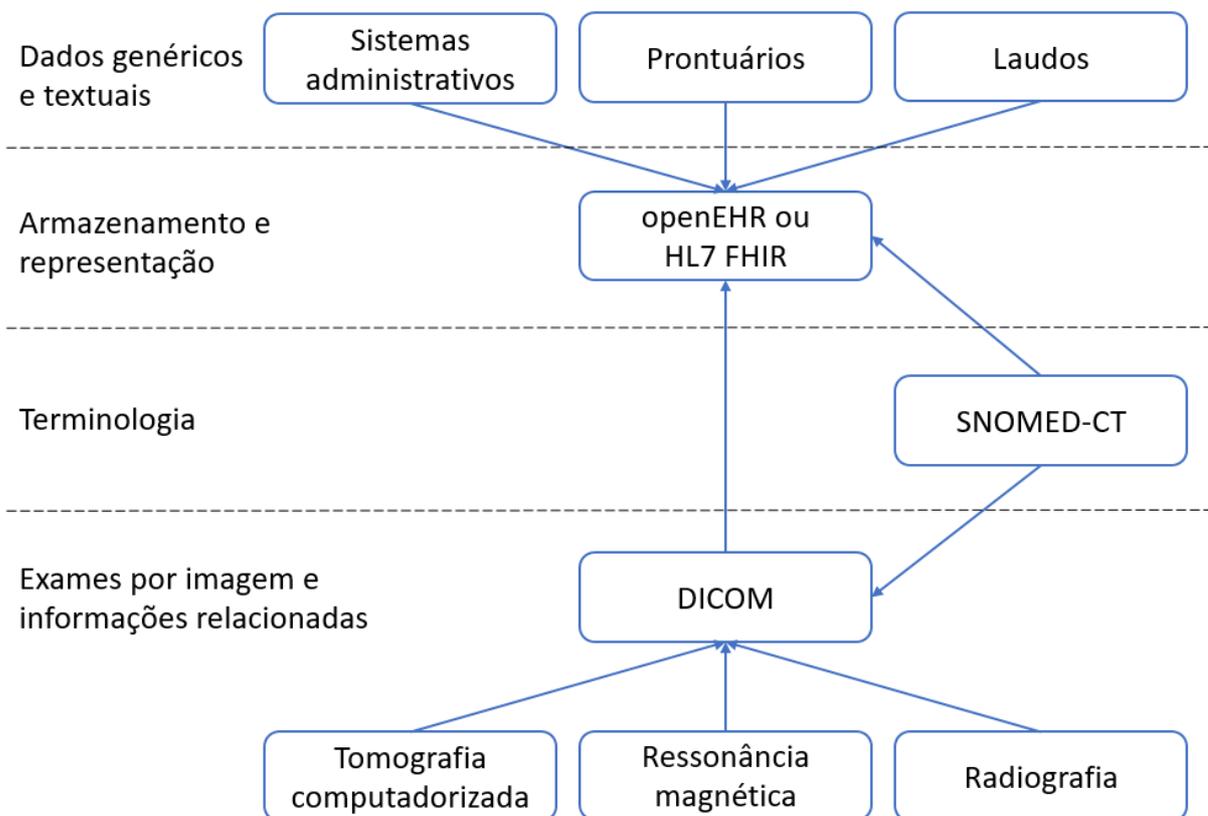
Todas as especificações de representação descritas permitem o emprego da terminologia SNOMED-CT. No HL7 FHIR, diversos elementos devem, obrigatoriamente, ser preenchidos de acordo com o sistema de codificação definido no SNOMED-CT. O mesmo existe no openEHR, em que os conceitos dessa terminologia podem ser utilizados para representar informações nas instâncias dos arquétipos.

O padrão DICOM também é empregado nas estruturas utilizadas no openEHR e no HL7, de modo que referências às imagens nesse formato podem ser inseridas nos arquétipos (como o *openEHR-EHR-CLUSTER.imaging.v1*) e nos recursos (como o *ImagingStudy*). Uma

vez que é possível armazenar informações sobre o equipamento, paciente, profissionais envolvidos, entre outros dados nos arquivos DICOM, Wang, Yao e Liu (2011) propuseram uma maneira de converter um arquivo nesse padrão para um arquétipo do openEHR, estruturando assim todos os dados relevantes armazenados em conjunto com a imagem. Com isso, passa a ser possível empregar esses dados diretamente em sistemas openEHR, simplificando assim sua utilização.

As especificações openEHR e HL7 FHIR são utilizadas com finalidades semelhantes, uma vez que ambas abordam o armazenamento estruturado e a troca de informações clínicas através do emprego de estruturas de dados definidas. A principal diferença é que os arquétipos do openEHR são conjuntos de informações que representam todo o conteúdo clínico, enquanto os recursos do HL7 FHIR contêm apenas os dados comumente utilizados, podendo ser estendidos com itens adicionais conforme a necessidade de uso (BOSCA et al., 2015). Segundo Allwell-Brown (2016), estes são os padrões que especificam a persistência e intercâmbio de dados clínicos mais recentes, robustos e completos, capazes de lidar com conjuntos de informações complexas, apoiando simultaneamente a interoperabilidade semântica. Pode-se determinar equivalências entre os recursos do HL7 FHIR e os arquétipos do openEHR, possibilitando alcançar certos níveis de integração entre os padrões. Segundo Bosca et al., 2015, acredita-se que esforços com o objetivo de possibilitar essa interoperabilidade serão cada vez mais frequentes. A Figura 18 ilustra as relações entre os tipos de dados e os padrões apresentados.

**Figura 18: Relações entre os tipos de dados e os padrões apresentados.**



Fonte: Elaborado pelo autor.



### 3 TRABALHOS RELACIONADOS

Neste capítulo são detalhados trabalhos que tem relação com o modelo proposto, selecionados a partir de pesquisas feitas nos portais *ScienceDirect*<sup>11</sup>, *ACM Digital Library*<sup>12</sup>, *Google Scholar*<sup>13</sup> e *PubMed*<sup>14</sup>, que agrupam revistas reconhecidas cientificamente e outros repositórios digitais. Foram avaliados apenas os resultados publicados nos últimos dez anos, disponíveis em periódicos científicos qualificados e eventos. Foi empregada a seguinte *string* de busca em cada um dos portais indicados: “(medical OR clinical OR medicine OR healthcare) AND (heterogeneous OR multimodal OR multi-source OR multimedia) AND (data OR image OR text OR report OR video OR sound OR sensor)”.

Através da leitura do título e do resumo dos resultados retornados foi possível identificar aqueles que não abordavam o uso de dados relacionados à área médica. Estes trabalhos não foram considerados, restando assim trinta e oito artigos. Estes, por sua vez, foram lidos com o objetivo de determinar sua aderência com a proposta do presente trabalho, ou seja, se fazem uso efetivo de diferentes tipos de informações clínicas, integrando as características obtidas em cada entrada e fornecendo resultados aos médicos ou especialistas. Esse é o principal filtro utilizado na seleção dos trabalhos relacionados, seguindo a definição de sistema multimídia de Riegler et al. (2016): sistema interativo, que provê suporte para a execução de diagnósticos, exames, cirurgias, construção de laudos e relatórios, além de ferramentas educacionais, que combinam todas as fontes de informações disponíveis, propiciando o acesso simplificado para os profissionais médicos e os pacientes. Do grupo citado, apenas seis trabalhos foram considerados como relacionados.

O objetivo dessa ação foi determinar as bases para os estudos preliminares do estado-da-arte em sistemas médicos que exploram diferentes fontes de dados em processos de apoio. Os projetos selecionados são apresentados nas seções seguintes.

#### 3.1 A modular framework for clinical decision support systems: medical device plug-and-play is critical

Nesta pesquisa, Williams et al. (2009) propuseram um *framework* modular para o desenvolvimento de sistemas de suporte à decisão clínica (CDSS), com ênfase na necessidade de se utilizar padrões *plug-and-play*<sup>15</sup>. O conceito básico está relacionado à definição de uma estrutura de *software* que seja responsável pela manipulação dos dados, gerenciamento e aspectos de apresentação. A aquisição e fusão das informações médicas, obtidas a partir de dispositivos que ficam ao lado dos leitos dos pacientes (medidores de frequência cardíaca, pressão arterial, saturação de oxigênio, dióxido de carbono, ventiladores e bombas de

---

<sup>11</sup> *ScienceDirect* – <http://www.sciencedirect.com/>

<sup>12</sup> *ACM Digital Library* – <http://dl.acm.org/>

<sup>13</sup> *Google Scholar* – <https://scholar.google.com/>

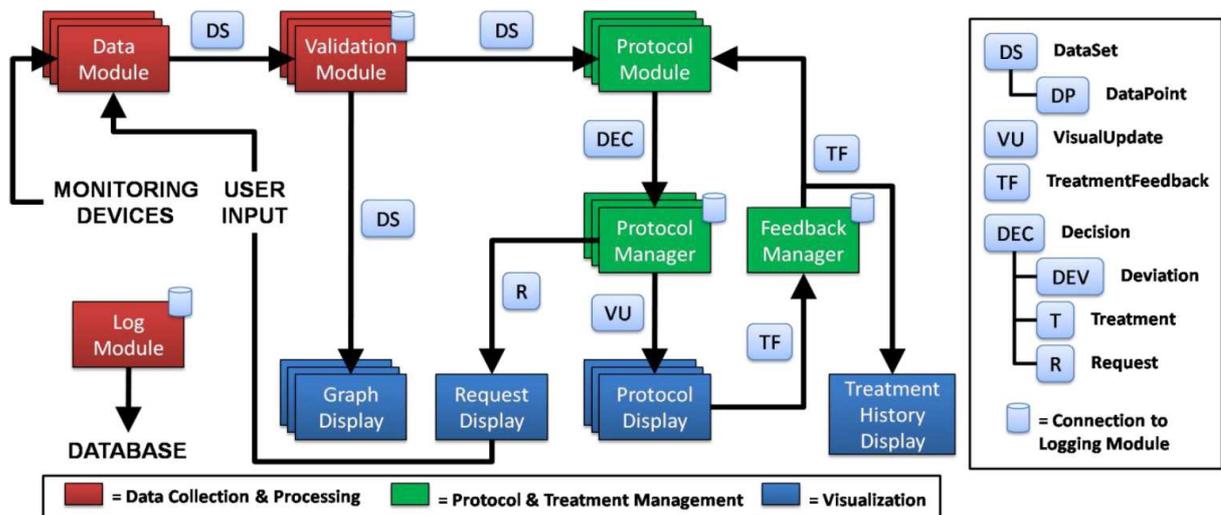
<sup>14</sup> *PubMed* – <https://www.ncbi.nlm.nih.gov/pubmed/>

<sup>15</sup> *Plug-and-play* – Adição de um novo componente que é automaticamente reconhecido em um sistema, passando a funcionar sem necessidade de configuração.

infusão), bases de dados de registros dos pacientes e informações manuais inseridas pela equipe clínica, é um dos componentes principais do *framework*.

O *framework* utiliza um design modular, que possibilita expansão e escalabilidade. Os componentes podem ser agrupados em três diferentes áreas: obtenção e processamento de dados; gerenciamento de protocolos e tratamentos; e visualização. A Figura 19 apresenta os componentes do *framework* e seus respectivos agrupamentos. O módulo de protocolo é encarregado de analisar os dados e fornecer apoio ao diagnóstico. O fluxo do processo envolve a obtenção dos dados, processamento das informações e apresentação para o usuário, que pode visualizar as informações e interagir, retornando dados para o sistema (*feedback*).

Figura 19: Componentes e agrupamentos do *framework* para CDSS.



Fonte: Williams et al. (2009).

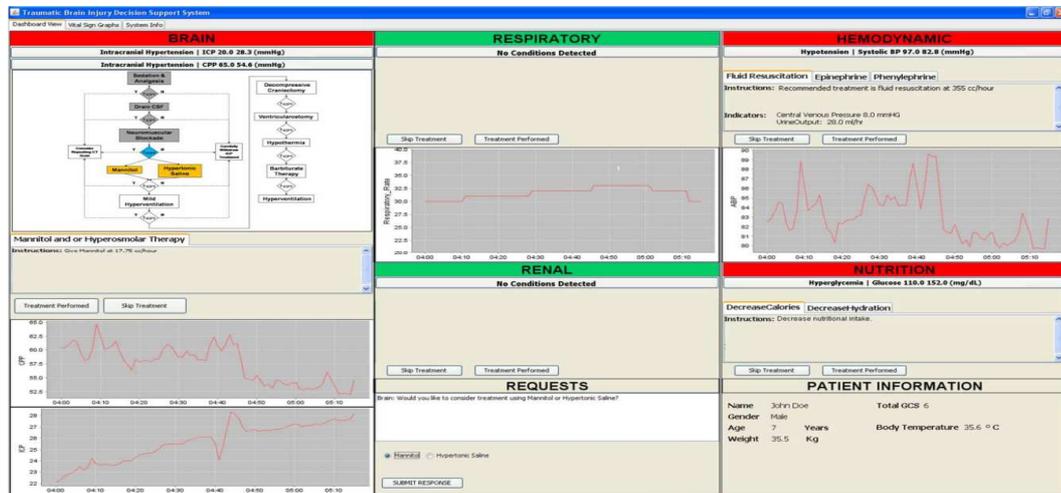
O tratamento dos dados é feito de forma assíncrona, com o objetivo de possibilitar o uso e visualização das informações em tempo real. A principal estrutura de dados utilizada é o *DataPoint*, que é composto por uma etiqueta de origem, data, hora, nome do dado e o valor correspondente. Um conjunto de *DataPoints* forma um *DataSet*. Outra estrutura de destaque é a denominada *Decision*, que contém os objetos *Deviation*, *Condition*, *Treatment* e *Request*. O primeiro é um alerta, utilizado para indicar quando uma variável controlada está fora do seu intervalo configurado. O segundo representa a condição clínica correspondente às variáveis que estão em *Deviation*. O terceiro indica qual procedimento deve ser tomado em resposta à condição. Por fim, em *Request* estão os questionamentos feitos ao usuário, que são importantes para que o sistema seja capaz de gerar recomendações.

As informações de sensores, condições do paciente, informações complementares e pontos de interação com o usuário são ilustrados na Figura 20. A aplicação criada para validar o *framework* é utilizada no acompanhamento e tratamento de pacientes com traumatismo craniano.

A principal dificuldade encontrada no desenvolvimento, aplicação e uso do *framework* proposto está relacionada ao fato de que não foram encontrados padrões universais de entrada e saída de informações dos dispositivos de monitoramento médico utilizados. Dessa forma, a característica *plug-and-play* é perdida, uma vez que a necessidade de se conectar um novo equipamento ao sistema envolve a criação de uma interface para comunicação. Seguindo o mesmo conceito, os autores apontaram dificuldades com a obtenção dos dados do paciente e de exames laboratoriais, pois essas informações também não se encontram armazenados de

maneira padronizada, reduzindo assim o número de ambientes em que o sistema pode ser empregado.

Figura 20: Interface de uso para validar o *framework plug-and-play*.



Fonte: Williams et al. (2009).

### 3.2 Redesigning EHRs and Clinical Decision Support Systems for the Precision Medicine Era

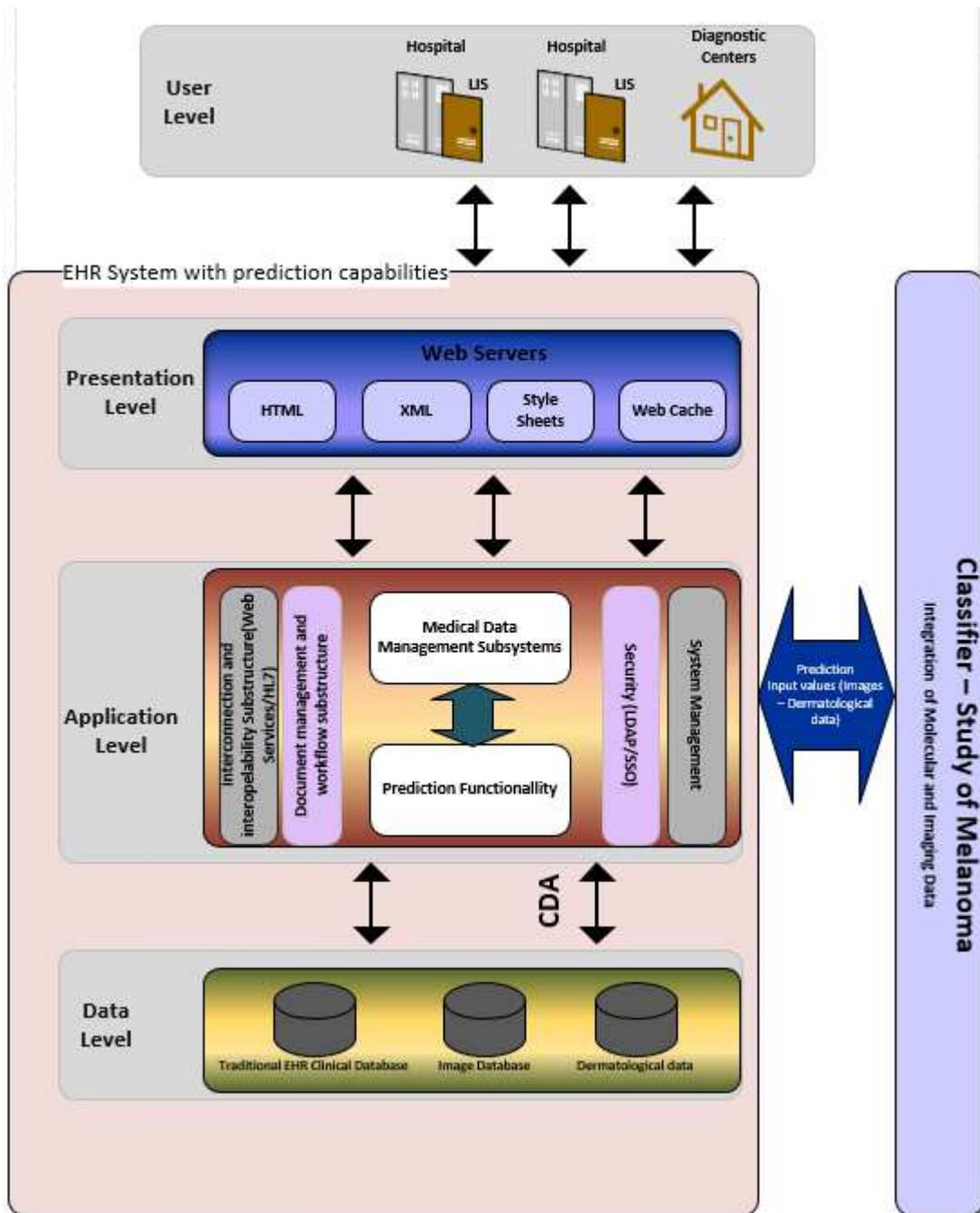
Neste trabalho, Maglogiannis et al. (2015) apresentaram uma arquitetura distribuída para sistemas de registro de saúde, capaz de integrar o processamento de informações heterogêneas de pacientes com melanoma<sup>16</sup>. Dentre os dados utilizados estão aspectos descritos no EHR, informações moleculares, observações clínicas e imagens das lesões de pele. A característica distribuída que o sistema possui está relacionada ao fato de que parte dos dados pode ser processado em diferentes locais, enquanto a outra parte é providenciada por um sistema centralizado. Como um dos objetivos é expandir o uso do EHR, são disponibilizadas características extras ao usuário, que não se encontram usualmente nesse tipo de sistema, como dados de exposição à radiação ultravioleta, informações hereditárias, regiões do corpo que foram afetadas, bem como recursos relacionados às lesões (assimetria, diâmetro, cor, alterações de tamanho). O sistema é composto por quatro camadas: base de dados, aplicação, apresentação e usuário, conforme apresentado na Figura 21.

A primeira camada é responsável por armazenar os dados operacionais e arquivos auxiliares. Além das informações tradicionais existentes no EHR, nesse local ficam inseridos também os dados dermatológicos e as imagens relacionadas. Na segunda camada estão os algoritmos referentes à segurança, gerenciamento do sistema, interconexão, interoperabilidade e predição, que é obtida a partir de um *webservice* para comunicação com o classificador externo, onde é feita a integração das informações heterogêneas. Já a camada de apresentação faz referência à interface de uso do sistema, onde são exibidas as informações de acordo com o papel e permissões de cada usuário. Para isso foram utilizadas tecnologias *web*, possibilitando a operação a partir de um navegador. Por fim, a camada de usuário é composta

<sup>16</sup> *Melanoma* – Tipo de câncer de pele.

pelas unidades públicas e privadas de saúde (hospitais, clínicas, médicos e outros profissionais), que podem estar conectados ao sistema.

Figura 21: Arquitetura do sistema que utiliza dados heterogêneos na predição de melanoma.



Fonte: Maglogiannis et al. (2015).

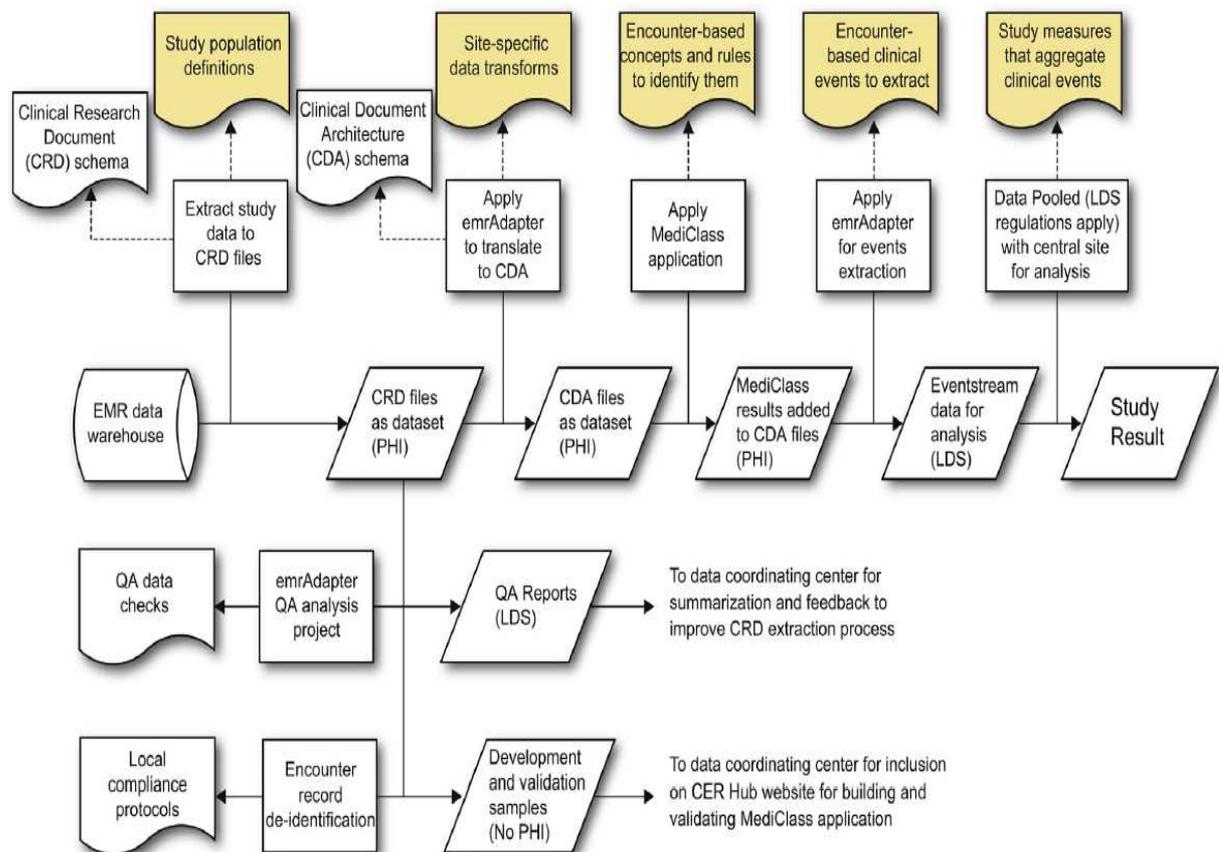
Segundo os autores, o uso do sistema proposto melhora a qualidade do serviço, proporcionando melhor acessibilidade. O uso de múltiplas fontes de informações possibilita diagnósticos com maior precisão, uma vez que são avaliadas mais características de cada paciente. A escolha por um sistema com apresentação via *web* possibilita o acesso à informação independentemente da localização, podendo ser disponibilizada para o próprio paciente. O foco principal da pesquisa foi aprimorar o EHR, a partir da inserção de

informações que auxiliam na predição de melanoma. Isso foi possível com a troca de dados entre diferentes sistemas de informações clínicas, demonstrando a possibilidade de interoperabilidade entre as instituições participantes.

### 3.3 CER Hub: An informatics platform for conducting comparative effectiveness research using multi-institutional, heterogeneous, electronic clinical data

Neste projeto, Hazlehurst et al. (2015) descreveram uma plataforma *web* para pesquisas de efetividade baseadas em comparação (*comparative effectiveness research*, ou CER), a partir da identificação, captura, agregação, integração e análise de dados localizados em EHRs de diferentes instituições, além de outras fontes de informações textuais. Um dos pré-requisitos para determinar o funcionamento apropriado de um sistema CER é a utilização de dados provenientes de várias organizações, de modo que existam informações suficientes para identificar diferenças estatisticamente significativas entre os grupos, melhorando a generalização dos resultados do estudo e permitindo a identificação de eventos raros. Essa tarefa é complexa de ser realizada, uma vez que organizações distintas se referem, com frequência, à uma mesma atividade, condição ou procedimento empregando termos diferentes, que em muitos casos possuem significados divergentes.

Figura 22: Fluxo de processamento da plataforma CER Hub.



Fonte: Hazlehurst et al. (2015).

O fluxo de processamento da plataforma é exibido na Figura 22. Em amarelo, no topo, estão as atividades que fazem referência a um estudo específico, que necessita de dados que não são utilizados de forma abrangente. Os símbolos em branco representam as diferentes transformações que são feitas nas informações, desde a obtenção dos dados (*EMR data warehouse*, à esquerda) até a disponibilização do resultado, à direita. As transformações feitas envolvem operações de extração, modelagem, agregação e análise dos dados. A entrada da plataforma é um protocolo que define a população sobre a qual será aplicada o estudo, as perguntas de pesquisa, os métodos de análise e a especificação das informações a serem utilizadas, como, por exemplo o período de aquisição. A operação de maior importância para o uso da plataforma é a definição das variáveis e dos eventos que devem ser avaliados, o que delimita os dados que serão extraídos e utilizados. A estrutura criada pelos autores contém informações como sinais vitais, medicações, procedimentos, diagnósticos, progresso do estado do paciente, descrições dos atendimentos, entre outros dados (exemplificados na Tabela 3). Essa estrutura é preenchida e representada com dados extraídos dos EHRs, a partir do padrão HL7-CDA.

**Tabela 3: Composição de parte da estrutura de dados utilizada no CER Hub.**

<b>Seção</b>	<b>Nº de variáveis</b>	<b>Exemplos de variáveis</b>
Detalhes do paciente	8	Gênero, raça, etnia
Detalhes do atendimento	6	Tipo, departamento
Provedores do atendimento	3	Tipo, departamento
Sinais vitais	12	Frequência cardíaca, pressão arterial, temperatura
Diagnóstico do atendimento	5	Código, padrão de codificação, ordem
Problemas	6	Código, padrão de codificação, estado
Medicações	25	Código, padrão de codificação, estado
Imunizações	5	Código, padrão de codificação
Alergias	7	Código, padrão de codificação, severidade, estado
Procedimentos	5	Código, padrão de codificação, estado
Notas de progresso	6	Código, padrão de codificação, estado
Recomendações	5	Código, padrão de codificação

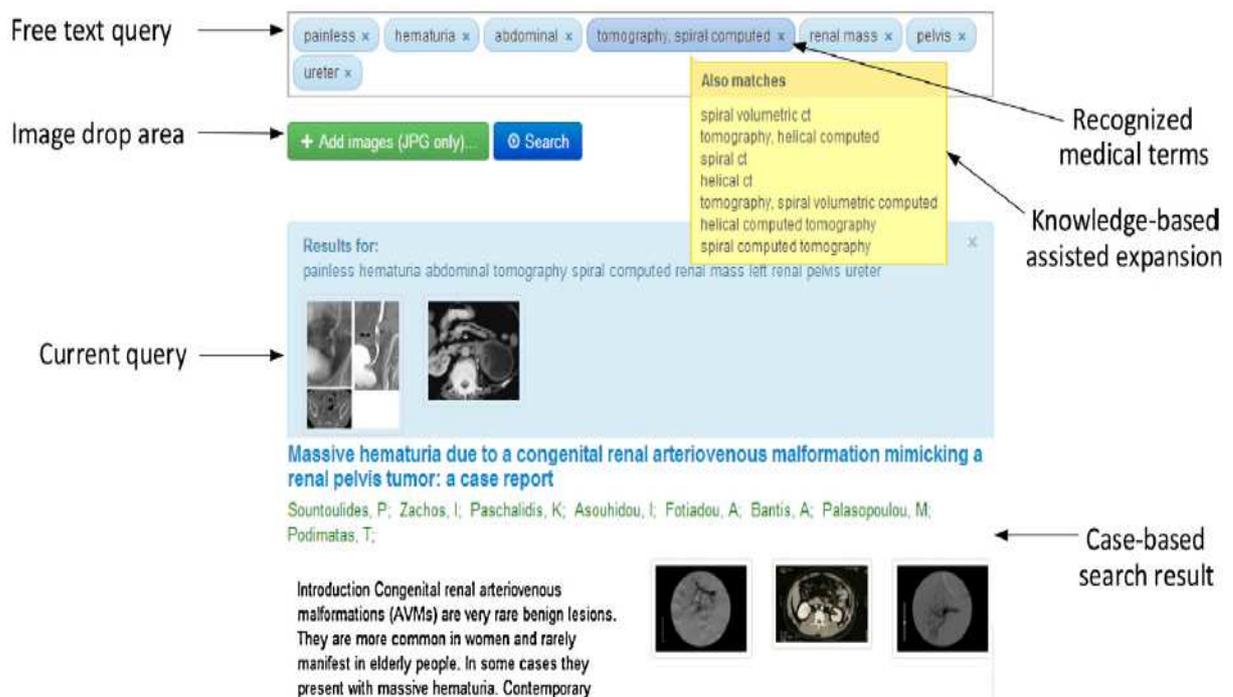
Fonte: Adaptado de Hazlehurst et al. (2015).

Além das informações padronizadas, são utilizadas ferramentas para extrair eventos de textos redigidos em linguagem natural, com a aplicação de técnicas de PLN e regras de classificação específicas para cada estudo. Os conceitos extraídos são representados com terminologias disponíveis no UMLS. O processamento das informações é feito de forma distribuída, onde cada instituição participante é responsável por analisar os dados da sua base. Isso evita preocupações referentes à segurança, uma vez que o acesso é feito de forma local, com permissões controladas e supervisão dos responsáveis. A saída não contém referências a um paciente específico, sendo composta pelos eventos clínicos que atenderam as condições definidas no protocolo de entrada. No total, seis instituições participaram do projeto, resultando em 2,5 milhões de pacientes. Os estudos feitos estão relacionados ao uso de medicamentos para controle de asma e serviços de apoio a tabagistas.

### 3.4 Multimodal medical information retrieval with unsupervised rank fusion

Neste estudo, Mourão, Martins e Magalhães (2014) detalharam um sistema de recuperação de informações médicas com suporte a buscas multimodais de casos clínicos, a partir da execução de um algoritmo de fusão de dados e sugestões de termos existentes em uma enciclopédia médica. Esse tipo de sistema é especialmente importante pois possibilita aos profissionais encontrar publicações relacionadas a determinada condição, bem como outros casos clínicos semelhantes existentes em repositórios médicos, que podem conter informações relevantes sobre os sintomas e tratamentos dos pacientes. Entretanto, os sistemas de busca comumente utilizados, como o PubMed<sup>17</sup>, empregam apenas buscas por palavras-chave, que devem estar presentes no resumo, título, categoria ou corpo do texto, fazendo que não sejam retornadas informações relevantes que tratam das imagens contidas nesses materiais. Isso ocorre porque as figuras não estão, necessariamente, sendo referenciadas no texto.

Figura 23: Interface de uso da busca multimodal com pesquisa e resultado.



Fonte: Adaptado de Mourão, Martins e Magalhães (2014).

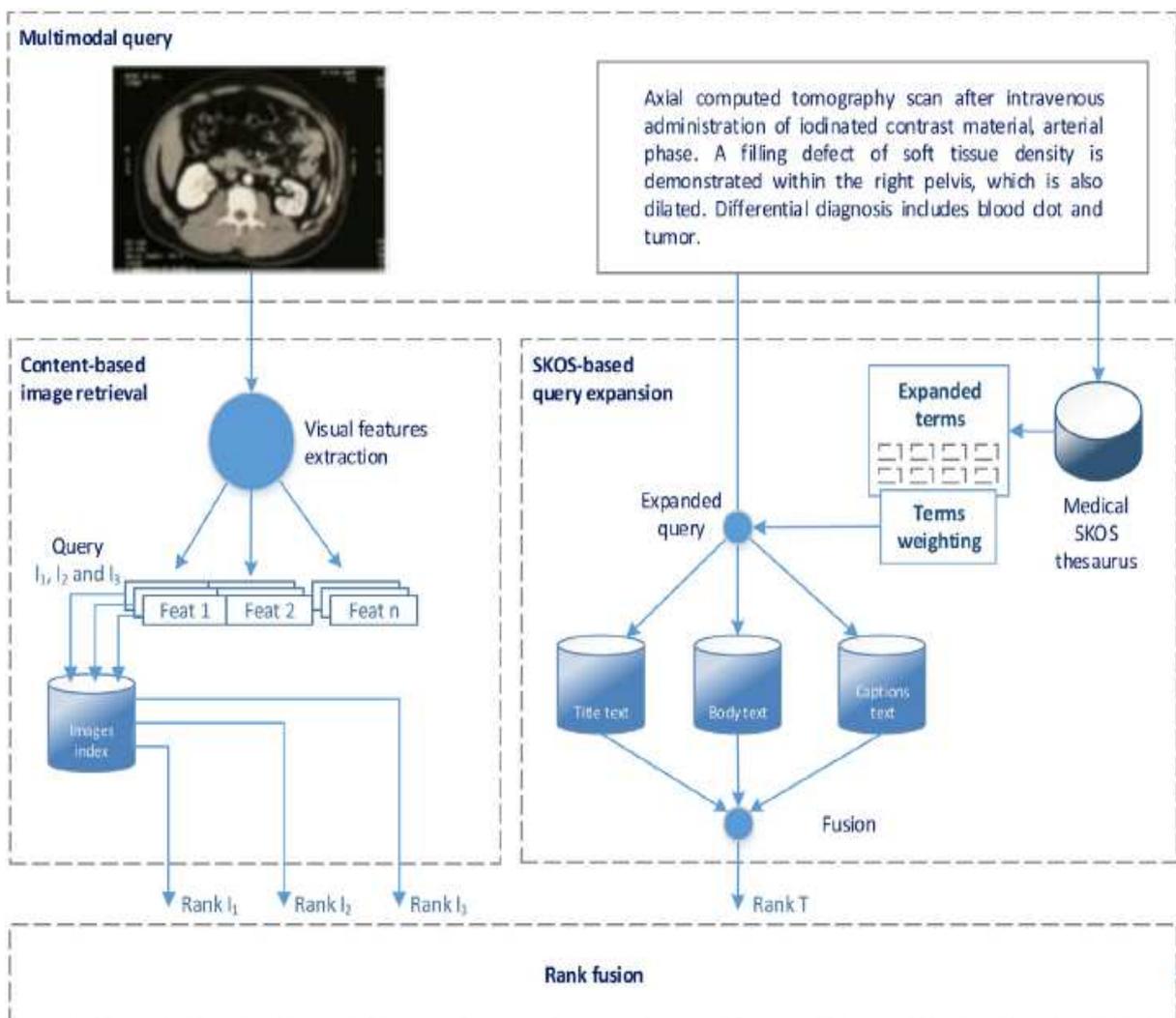
Com o objetivo de reduzir a perda na recuperação de dados importantes, os autores desenvolveram um sistema em que o especialista tem a possibilidade de escrever buscas textuais e complementar com imagens. Esses dois tipos de dados são combinados utilizando técnicas de fusão de informações, sendo então submetidos ao motor de busca. Outro recurso existente é a sugestão de expansão de termos, em que a sentença de pesquisa é ampliada automaticamente com palavras relacionadas às utilizadas na busca, aumentando a abrangência da consulta. Como a terminologia utilizada nos documentos médicos é composta por um grande número de definições e termos que fazem referência a um mesmo sintoma, medicação ou doença, a expansão automática impacta positivamente nos registros retornados em cada

<sup>17</sup> PubMed – <https://www.ncbi.nlm.nih.gov/pubmed/>

busca. A Figura 23 apresenta a interface de uso do sistema desenvolvido, cujo nome é *NovaMedSearch*.

O processamento aplicado sobre as imagens que podem ser pesquisadas consiste na divisão de cada figura em partes menores. Cada uma dessas partes é avaliada, e duas características são extraídas: histograma de cores HSV (*Hue Saturation Value*) e padrões binários locais (LBP). Esses atributos são concatenados, em conjunto com todos os atributos extraídos de todas as divisões da figura, formando assim um vetor de características, que é então inserido em um índice. O mesmo processo é feito nas imagens de entrada, possibilitando assim a execução de buscas pelas figuras similares que estão armazenadas no índice citado anteriormente. A Figura 24 ilustra a sequência descrita.

**Figura 24: Fluxo de funcionamento da busca multimodal.**



Fonte: Adaptado de Mourão, Martins e Magalhães (2014).

Já o processamento aplicado sobre as informações textuais é formado pela etapa de expansão das palavras utilizadas, de acordo com as suas definições e relações semânticas com outros termos. São utilizadas técnicas de PLN, como *tokenização*, *lematização* e *stemming* para aprimorar a eficácia da ação, já que essas operações reduzem o número de termos diferentes a serem avaliados. Após a expansão, os componentes da sentença de busca são

comparados com o conteúdo do título e corpo dos textos, além de eventuais legendas de imagens que podem existir. Algoritmos de fusão são utilizados para combinar os resultados encontrados e, posteriormente, determinar a ordem em que os resultados retornados são exibidos.

### **3.5 METEOR: An Enterprise Health Informatics Environment to Support Evidence-based Medicine**

Nesta pesquisa, Puppala et al. (2015) apresentaram o desenvolvimento de um ambiente computacional integrado, denominado METEOR (*Methodist Environment for Translational Enhancement and Outcomes Research*). Esse ambiente é composto por dois componentes: o *enterprise data warehouse* (EDW) e um software de inteligência e análise (SIA), para suporte ao exercício da medicina baseada em evidência e outros CDSS. O projeto foi motivado pelo fato de que não existe um sistema centralizado de acesso a informação que agrupe os múltiplos tipos de dados existentes nos sete hospitais que fazem parte do grupo *Houston Methodist Hospital*, tornando o processo de extração de dados para uso em pesquisas, consultas e testes de hipóteses uma tarefa custosa. O conjunto de informações existentes no ambiente desenvolvido é formado por registros de mais de dois milhões de pacientes, totalizando mais de dez milhões de registros de atendimentos. Dentre os dados que podem ser utilizados estão imagens, testes de laboratório, tratamentos, informações financeiras e administrativas.

O primeiro componente, denominado EDW, é baseado no sistema i2b2 (*Informatics for Integrating Biology and the Bedside*), com características expandidas e interfaces de busca criadas para cada tipo de usuário que utiliza o ambiente. Foi necessário definir quais informações eram necessárias, em que hospitais estavam localizadas, como estavam armazenadas e quais métodos seriam utilizados para acessá-las. A partir dos registros extraídos, foi formado um repositório centralizado que contém os dados utilizados nas consultas e no componente SIA. A Figura 25 apresenta o fluxo das informações, exibindo as conexões entre as diferentes fontes de dados, com o EDW representado no centro da imagem.

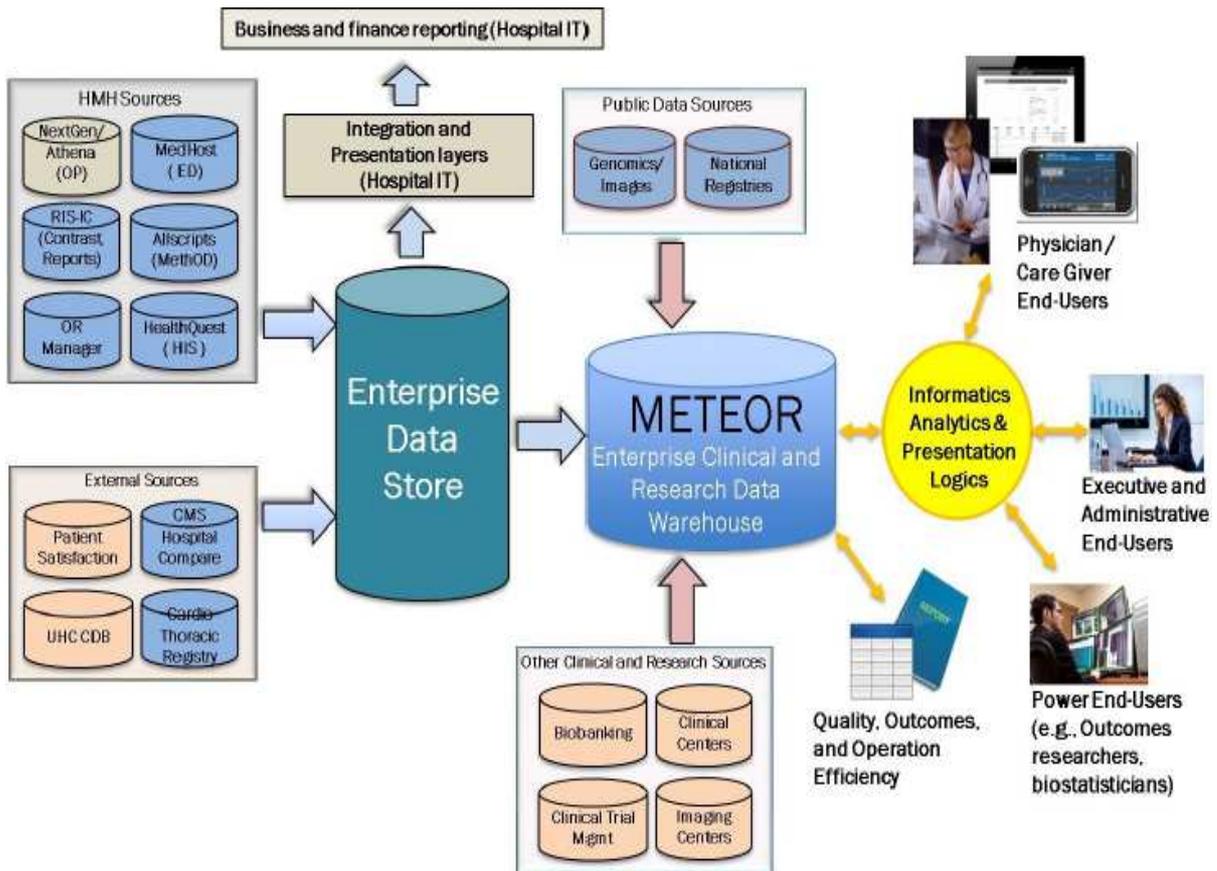
Na camada de inteligência e análise estão os portais de acesso *web* e *desktop*, interface com *smartphones*, monitoramento de sensores, sistemas de recuperação de informações, de alertas, de sugestões e modelos de predição. Além destes, com a integração de múltiplos tipos de dados é possível o uso de bioestatística e aprendizado de máquina em sistemas de suporte à decisão baseados em evidências. O trabalho cita três aplicações desenvolvidas na camada destacada: mPOD, MOTTE e READMIT.

A primeira, o mPOD, tem como finalidade a extração de dados de saúde a nível populacional, permitindo ao usuário, por exemplo, determinar grupos de pacientes que podem participar de testes clínicos ou estudos de determinadas condições. A interface de uso possibilita a definição de critérios de inclusão e critérios de exclusão, que podem estar relacionados à dados demográficos, detalhes das consultas, diagnósticos, procedimentos, dados laboratoriais e medicações. Esses critérios são selecionados a partir de listas pré-determinadas, de modo que o operador pode determinar os valores válidos para cada item.

A segunda é uma ferramenta para coleta e análise de dados que estão armazenados em textos não estruturados, com a aplicação de técnicas de processamento de linguagem natural em relatórios médicos, resultados de laboratórios e outros registros que foram redigidos manualmente pelos especialistas. No módulo de PLN são feitas operações de *tokenização*,

*stemming*, remoção de *stopwords* e modelagem de todos os *tokens* em um espaço vetorial de alta dimensionalidade. A partir do espaço vetorial modelado são feitos cálculos de similaridade com a questão de pesquisa utilizada como entrada da ferramenta. Foram extraídos parâmetros relacionados ao câncer de mama, possibilitando a definição de um escore de avaliação de risco. Esse valor é empregado pelo clínico para determinar, com maior precisão, se um paciente deve ou não ser submetido a uma biópsia, reduzindo assim os procedimentos falso-positivos.

**Figura 25: Fluxo de informações no ambiente METEOR.**



Fonte: Puppala et al. (2015).

Por fim, a última ferramenta desenvolvida é um aplicativo de *smartphone* que utiliza mineração de dados para prever pacientes com alto risco de readmissão, fazendo uso das informações médicas correspondentes (duração da estadia no hospital, número de condições identificadas na admissão, severidade, entre outras) em relação à casos semelhantes. A acurácia obtida ficou próxima de 70,9%.

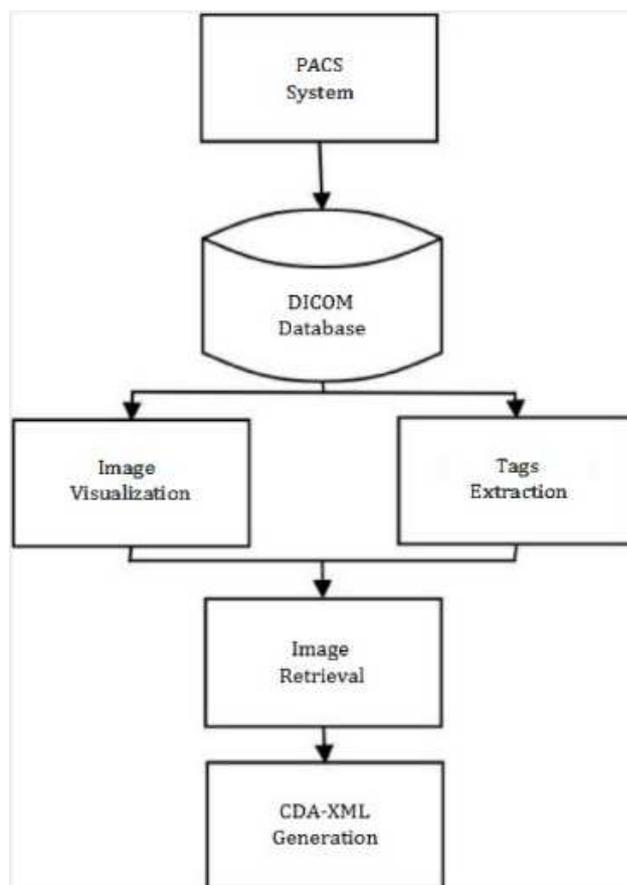
### 3.6 GIDAC: a prototype for bioimages annotation and clinical data integration

Neste trabalho, Vizza et al. (2016) descreveram um *framework* denominado GIDAC, que visa definir uma estrutura de propósito geral para o gerenciamento de imagens e suas respectivas anotações. Além destas, o modelo permite a visualização de dados clínicos, pesquisa e integração de diferentes informações, como histórico do paciente e dados

laboratoriais, em um único sistema de informação. Isso possibilita a identificação de casos de câncer semelhantes e os resultados obtidos a partir da aplicação de diferentes procedimentos. Atualmente, segundo os autores, a gestão desses dados é realizada por sistemas individuais dentro de cada departamento hospitalar, não estando disponíveis plataformas dedicadas à sua integração. Além disso, as imagens não podem ser anotadas ou enriquecidas por médicos para traçar estudos temporais ou mesmo entre pacientes com doenças semelhantes.

O sistema é composto por um conjunto de módulos, dentre os quais estão: estatístico, comunicação, terapias, suporte ao diagnóstico, integração de dados instrumentais e elaboração de imagens. O módulo estatístico contém a criação de listas e relatórios que podem ser usados pelas áreas administrativas na avaliação de tendências, que são extraídas a partir dos dados disponíveis. O módulo de comunicação trata do acesso às informações clínicas pelos médicos e especialistas, tanto em operações de simples consulta como na integração com outros módulos. Já o módulo de terapias contém as aplicações de gerência dos pacientes e seus respectivos tratamentos, com informações de diagnóstico, medicamentos e recursos necessários. O módulo de suporte ao diagnóstico é usado na definição de relações entre o diagnóstico e os relatórios médicos. O módulo de integração é o encarregado por simplificar o acesso aos diferentes tipos de dados utilizados no *framework*, como informações clínicas e pessoais dos pacientes, migrando características contidas nas imagens para textos. Por fim, o módulo de elaboração de imagens permite a visualização e manipulação das imagens, possibilitando, por exemplo, a inserção de anotações.

**Figura 26: Fluxo da geração de CDA-XML a partir das imagens.**



Fonte: Vizza et al. (2016).

O módulo de integração utiliza um modelo para relacionar os exames por imagem, que estão no padrão DICOM, com o padrão CDA XML, a partir do uso da biblioteca HAPI<sup>18</sup>. As imagens são extraídas do servidor PACS (*Picture Archiving and Communication System*) e agrupadas de acordo com o modo de aquisição e orientação. Posteriormente, as informações extraídas de cada imagem são codificadas e convertidas para o formato XML/HL7 com uso de CDA, possibilitando assim a interoperabilidade com os dados textuais e outros sistemas. O fluxo desse processo é apresentado na Figura 26.

No módulo de elaboração das imagens são empregadas, em sequência, uma série de métodos de processamento de imagens com o objetivo de identificar tumores, de acordo com a seguinte ordem: aquisição (tomografia computadorizada, ressonância magnética e tomografia por emissão de pósitrons), pré-processamento (redução de ruídos e suavização), definição das regiões de interesse, segmentação (limiarização e crescimento de regiões), extração de características (formato, calcificação e dimensões), classificação e predição.

### 3.7 Análise comparativa e lacunas de pesquisa

A partir da análise dos trabalhos relacionados, descritos nas seções anteriores, foi possível determinar as principais características existentes em cada um. Os critérios de análise utilizados para classificá-los foram definidos a partir das estruturas propostas nos trabalhos, resultando em quatro grandes grupos: entrada, extração de características, padrões de representação e aplicações. Estes, por sua vez, foram divididos em subgrupos que especificam cada um dos seus itens. Nos critérios de entrada e extração de características, as subdivisões estão relacionadas aos tipos de dados. Já no item que aborda os padrões de representação, as subdivisões são: padrões e uso de bases externas. Abaixo estão descritos os elementos avaliados:

- a) Entrada: Indica quais tipos de dados o sistema utiliza como entrada;
- b) Extração de características: Indica se o sistema possui, ou possibilita, métodos para extrair características de cada um dos tipos de dados de entrada;
- c) Padrões: Aponta os padrões utilizados para armazenar e representar o conhecimento extraído das entradas do sistema. Nesse grupo também é avaliado se o sistema possibilita o uso de informações complementares, a partir de integrações com bases de conhecimento externas, ou se os dados existentes no sistema são disponibilizados para aplicações de terceiros;
- d) Aplicações: Cita as diferentes aplicações desenvolvidas que fazem uso das informações extraídas a partir das entradas, indicando assim a finalidade de cada um dos sistemas.

Na Tabela 4 constam os itens analisados, os trabalhos relacionados e suas respectivas classificações. É possível verificar que todos utilizam mais de uma fonte de informações, mas nenhum deles apresenta métodos de processamento de mais de dois tipos de dados. Em Hazlehurst et al. (2015) são utilizados apenas conceitos na forma de textos, que podem estar representados de maneira estruturada ou não estruturada. Todos os sistemas manipulam informações textuais, obtidas através de registros clínicos e EHRs. Somente a pesquisa de

---

<sup>18</sup> HAPI – <http://hl7api.sourceforge.net/>

Williams et al. (2009) apresentou integração direta com sensores. Os demais utilizam, além de textos, imagens médicas como entrada. O emprego de vídeos e sons não foi explorado, entretanto, conforme apresentado por Pampouchidou et al. (2016), esse tipo de informação pode contribuir na identificação de condições e na tomada de decisão.

A primeira lacuna identificada está relacionada à extração e integração de informações a partir do processamento de diversos tipos de entradas, como textos, imagens, vídeos, sons e sensores. Essas ações, segundo Weese e Lorenz (2016), podem contribuir para a obtenção de melhores resultados em sistemas de apoio, além de possibilitar o desenvolvimento de novas aplicações.

**Tabela 4: Classificação dos trabalhos relacionados quanto às suas características.**

Critérios		Trabalhos relacionados					
		Williams et al. (2009)	Maglogiannis et al. (2015)	Hazlehurst et al. (2015)	Mourão, Martins e Magalhães (2014)	Puppala et al. (2015)	Vizza et al. (2016)
Entrada	Textos	Sim	Sim	Sim	Sim	Sim	Sim
	Imagens	Não	Sim	Não	Sim	Sim	Sim
	Vídeos	Não	Não	Não	Não	Não	Não
	Sons	Não	Não	Não	Não	Não	Não
	Sensores	Sim	Não	Não	Não	Não	Não
Extração de características	Textos	Sim	Sim	Sim	Sim	Sim	Sim
	Imagens	Não	Sim	Não	Sim	Não	Sim
	Vídeos	Não	Não	Não	Não	Não	Não
	Sons	Não	Não	Não	Não	Não	Não
	Sensores	Sim	Não	Não	Não	Não	Não
Padrões de representação	Padrões	Próprio	Não descrito	-HL7 CDA -UMLS -Próprio (CRD)	Não descrito	Extensão de i2b2	- DICOM - XML - HL7 CDA - ICD9
	Uso de bases externas	Não	Não	Não	Sim	Sim	Não
Aplicações		Acompanhamento e tratamento de pacientes com traumatismo craniano	Predição de melanoma	- Eficácia de medicamentos para asma. - Eficácia de apoio a tabagistas.	Busca multimodal de casos clínicos	- Lista de candidatos para testes clínicos. - Risco de readmissão.	Anotação de imagens médicas

Fonte: Elaborado pelo autor.

Outra lacuna percebida na Tabela 4 é a falta de um padrão comum, adotado em larga escala, para a representação e o armazenamento dos dados obtidos a partir do processamento das entradas. Esse aspecto é especialmente importante, pois permite a interoperabilidade entre diferentes sistemas, uma vez que as informações extraídas estarão descritas de uma maneira que pode ser reconhecida por outras aplicações que utilizam o mesmo padrão especificado, permitindo o compartilhamento entre elas. Por essa razão, o modelo proposto visa simplificar

a adoção de padrões, apoiando a popularização do seu uso. Outra possibilidade é a utilização de dados disponibilizados por terceiros, que também podem estar acessíveis de acordo com o padrão escolhido. Duas pesquisas fazem uso do padrão HL7 CDA. As demais não descrevem o esquema utilizado ou empregam um formato próprio para representação e armazenamento dos dados.

Alguns dos trabalhos indicam que os respectivos protótipos desenvolvidos são expansíveis, permitindo ampliar seu volume de funcionalidades. Porém, não foram detalhados exemplos de uso dessa característica, além de não terem sido citados trabalhos futuros que pudessem explorar essa propriedade, ou mesmo maneiras para serem utilizados métodos já desenvolvidos por outros pesquisadores. A falta de um modelo que permita agregar novos recursos de maneira simples é considerada a terceira lacuna identificada. A área médica é bastante ampla, de modo que a possibilidade de incrementar os sistemas de apoio clínico com o desenvolvimento de novas aplicações, que fazem uso dos dados já existentes, é um requisito desejável.

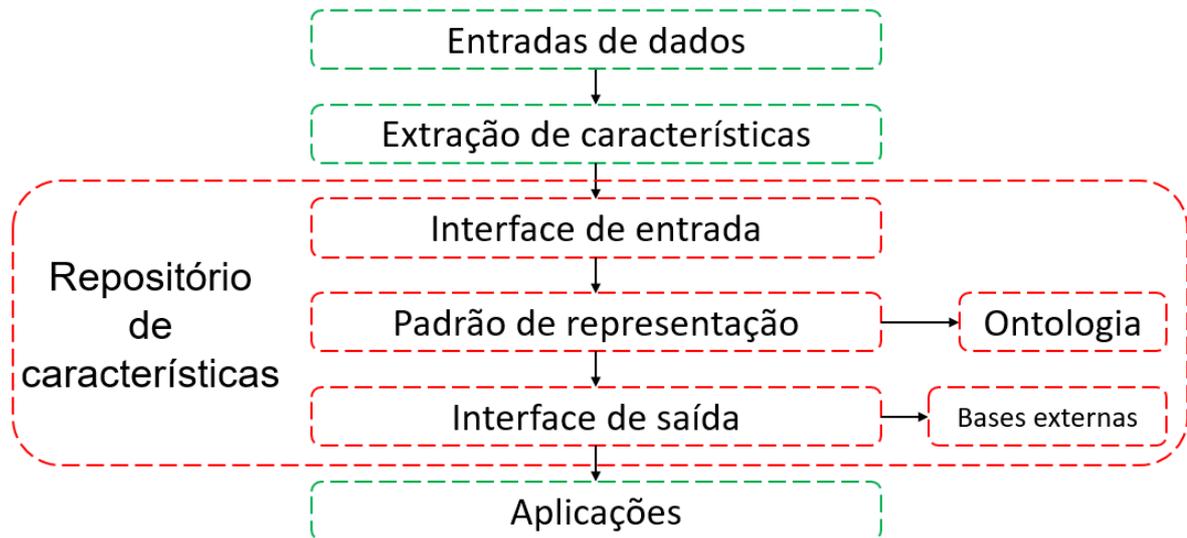
Os três aspectos citados como lacunas estão diretamente relacionados à questão de pesquisa descrita anteriormente. A primeira lacuna é relativa ao uso integrado de informações extraídas em diferentes tipos de dados. Já a segunda e a terceira estão ligadas à interoperabilidade com sistemas já desenvolvidos. O preenchimento de cada uma das lacunas fomenta os critérios de aceitação da solução que o modelo proposto busca atender, com ênfase para o uso de dados provenientes de múltiplas fontes e para a possibilidade de adição de funcionalidades a partir do emprego de pesquisas já publicadas. Essas lacunas compõem a base de especificações do modelo proposto neste trabalho, que é detalhado no capítulo seguinte.

## 4 MODELO PROPOSTO

O principal objetivo deste modelo é apresentar uma estrutura que possa ser utilizada como base para a construção de aplicações de apoio clínico com diferentes finalidades, viabilizando a utilização de diferentes tipos de dados de forma integrada. Este modelo define a composição e a organização de aplicações desse tipo, viabilizando o emprego e a interoperabilidade com sistemas já desenvolvidos, ampliando assim a contribuição da computação e auxiliando os profissionais da saúde na execução de suas tarefas. Após o estudo da bibliografia e da análise dos trabalhos relacionados foi possível determinar os elementos que o modelo deve conter, bem como o relacionamento entre eles, de modo que seja possível alcançar o objetivo citado.

O modelo é dividido em quatro camadas: entradas de dados, extração de características, repositório de características e, por fim, aplicações. A Figura 27 ilustra a estrutura geral do modelo, com destaque para a camada do repositório das características. O primeiro grupo representa os diferentes tipos de dados que podem ser utilizados como entrada. O segundo grupo contém os processos de extração de características correspondentes a cada tipo de entrada. O terceiro grupo, que é o principal diferencial do modelo proposto, contém os elementos utilizados para viabilizar o acesso, o armazenamento e a visualização das características extraídas a partir das informações utilizadas como entradas. O quarto e último grupo contém as aplicações que fazem uso integrado dos dados extraídos em atividades de suporte clínico.

Figura 27: Estrutura geral do modelo proposto.

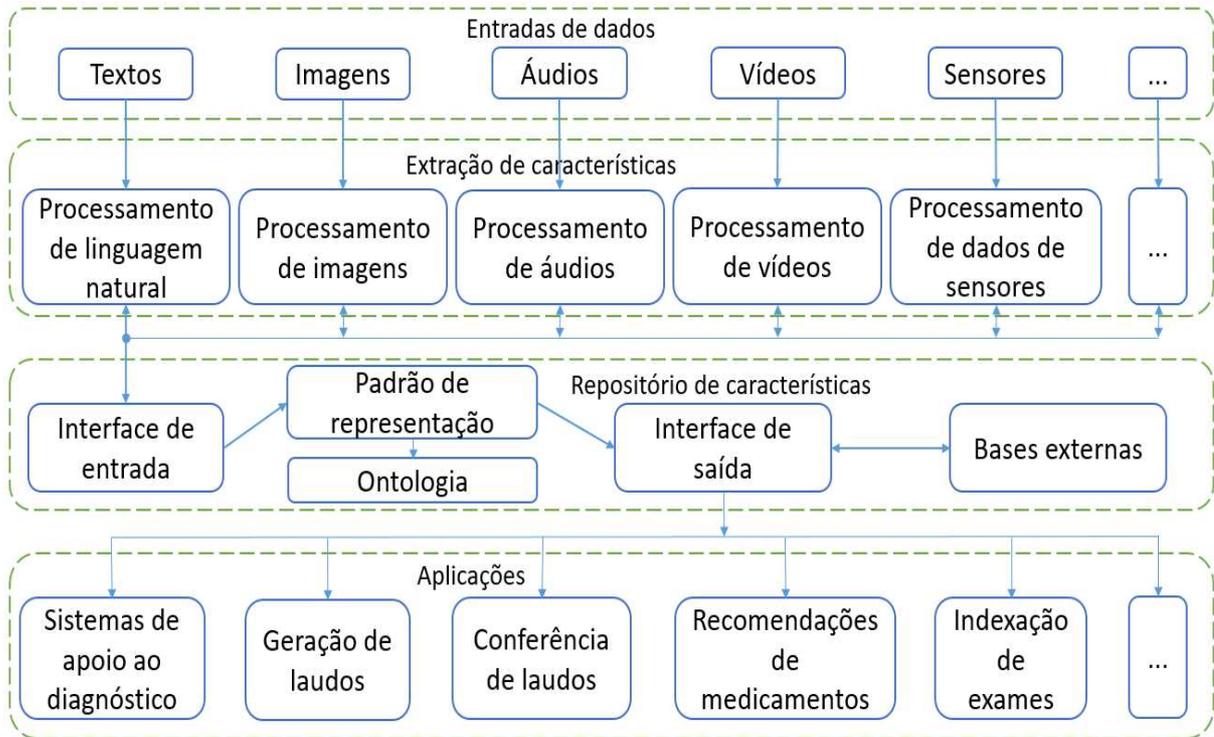


Fonte: Elaborado pelo autor.

Em cada uma das camadas apresentadas podem ser inseridas instâncias de elementos correspondentes, que especificam e implementam o uso de diferentes tipos de entradas, padrões de representação e aplicações. A Figura 28 apresenta alguns exemplos de componentes que podem compor os sistemas construídos a partir do modelo proposto. Toda a estrutura do modelo é organizada de maneira modular, ou seja, cada componente é separado e representado por um bloco, que está interligado e relacionado aos demais. O objetivo é

possibilitar a adição de funcionalidades conforme o desenvolvimento e necessidade de novos recursos, ampliando a sua abrangência.

**Figura 28: Exemplos de elementos que podem existir em cada camada do modelo proposto.**



Fonte: Elaborado pelo autor.

Outra característica desse formato de construção é a possibilidade de alterar um componente específico sem a necessidade de modificar os demais, uma vez que os módulos são independentes entre si. Por fim, também é possível empregar um mesmo bloco em diferentes operações, de modo que as funções já desenvolvidas possam ser reaproveitadas. Nas seções seguintes são detalhados cada uma das camadas, bem como os exemplos de elementos que podem as compor.

#### 4.1 Entradas de dados

Os elementos que compõem o grupo denominado Entradas de dados representam os diferentes tipos de dados que podem ser utilizados como fontes para obtenção de características. Informações textuais (BOZKURT et al., 2016; PONS et al., 2016; TOPAZ et al., 2016; HAZLEHURST et al., 2015) e imagens (CHENG et al., 2016; EL-DAHSHAN et al., 2014; FIRMINO et al., 2016) são as categorias utilizadas com maior frequência em sistemas de apoio à tomada de decisão na área médica. Além destas, já foram publicadas pesquisas que apresentam o uso de vídeos (RIEGLER et al., 2016), áudios (PAMPOUCHIDOU et al., 2016) e dados de sensores (WILLIAMS et al., 2009) em aplicações com essa finalidade.

A possibilidade de ser submetido, como entrada do modelo, qualquer tipo de dado disponível na área da saúde visa atender a primeira lacuna de pesquisa, identificada a partir da

avaliação dos trabalhos relacionados, que trata da utilização de informações provenientes de fontes heterogêneas. Na Figura 28 foram citados alguns exemplos, como textos, imagens, áudios, vídeos e sensores. Cada um desses grupos pode ser expandido em diferentes subgrupos. Por exemplo, informações textuais são encontradas de maneira estruturada ou na forma de textos escritos em linguagem natural, com o emprego ou não de terminologias específicas. Sendo assim, de acordo com os formatos de entrada, podem ser necessárias diferentes implementações relacionadas ao uso de textos clínicos. O mesmo ocorre no uso de imagens médicas, que são usualmente representadas através do padrão DICOM. Entretanto, também é praticável o emprego de outros padrões, como BMP<sup>19</sup>, JPEG<sup>20</sup> ou PNG<sup>21</sup>. Nesse sentido, podem existir elementos distintos na camada denominada Entrada de dados, que correspondem a cada um dos tipos de imagens que foram citados.

Devido à característica modular, caso seja identificada a necessidade de utilizar um formato de dados diferente dos já especificados, é possível adicioná-lo no modelo sem que seja preciso modificar os demais elementos relacionados às entradas. Essa possibilidade atende a terceira lacuna indicada, viabilizando a ampliação do número e do formato das informações que podem ser submetidas ao sistema.

## 4.2 Extração de características

Cada tipo de dado empregado como entrada é processado por uma rotina correspondente, que tem como finalidade extrair as informações que poderão ser utilizadas, posteriormente, pelas aplicações. Os processos responsáveis por essa tarefa estão contidos na camada denominada Extração de características. A obtenção das características é feita de acordo com o tipo da entrada, através de novos algoritmos ou métodos apresentados em pesquisas já divulgadas. Nos protótipos descritos no próximo capítulo estão detalhados os elementos desenvolvidos nessa camada, responsáveis pelo processamento de imagens de exames de tomografia computadorizada, de textos dos laudos correspondentes aos exames e de dados provenientes de sensores de temperatura corporal, pressão sanguínea e frequência cardíaca.

É possível utilizar, nessa camada, um ou mais componentes para cada tipo de dado, de acordo com os padrões utilizados na entrada ou das informações que se deseja obter. Conforme citado na seção anterior, por exemplo, podem ser utilizadas imagens médicas que estão em diferentes formatos. Consequentemente, podem existir diferentes implementações para cada padrão de entrada, mesmo que todas sejam utilizadas para avaliar imagens. Além do formato, é prevista a existência de rotinas distintas que são aplicadas no processamento de uma mesma entrada, mas que tem como finalidade a obtenção de informações específicas.

Um caso de uso que apresenta o emprego de dois métodos diferentes em uma mesma entrada, mas para a obtenção de informações distintas, é o uso de imagens de tomografia computadorizada do tórax para a detecção, nos pulmões, de nódulos e de enfisema. Em relação à detecção de nódulos nessa região, é possível o emprego do método descrito por Shaukat et al. (2017). Já a identificação de enfisema nesse local pode ser alcançada através do

---

<sup>19</sup> *BMP – Bitmap*

<sup>20</sup> *JPEG – Joint Photographic Experts Group*

<sup>21</sup> *PNG – Portable Network Graphics*

algoritmo proposto por Ginsburg et al. (2012). Essa forma de uso apresenta dois componentes que atuam sobre uma mesma entrada, extraindo informações diferentes.

Na literatura é possível encontrar diversas técnicas para extração de características a partir do processamento de cada um dos tipos de entradas citados. Nabih-Ali, El-Dahshan e Yahia (2017) agruparam, em seu estudo, o estado da arte em métodos de análise de sinais sonoros cardíacos. Já Jaiswal e Banka (2017) descreveram uma abordagem automatizada para identificar a ocorrência de episódios de convulsão a partir da análise dos sinais de exames de eletrocardiograma.

Como pode ser observado a partir das pesquisas citadas anteriormente, que fazem referência ao uso de imagens, de sons e de dados provenientes de sensores, o processamento de cada tipo de entrada é feito de forma individual, não havendo a necessidade de existir um único componente capaz de processar qualquer informação e extrair qualquer característica. A viabilidade de serem adicionados ao sistema diferentes elementos, cada um sendo responsável pelo processamento de um ou mais tipos de informações, é um exemplo do emprego de pesquisas já desenvolvidas e validadas, além de reforçar a capacidade modular e expansível do modelo. A possibilidade de adicionar novos algoritmos para obtenção de informações complementa a terceira lacuna de pesquisa citada, que trata da ampliação de funcionalidades do sistema sem a necessidade de se modificar os demais componentes já utilizados.

As rotinas de extração têm, como saída, as características que serão armazenadas, de maneira estruturada, no repositório de características. Essa ação complementa o aspecto heterogêneo do modelo (primeira lacuna de pesquisa), já que é uma etapa fundamental para o uso, a partir das aplicações, dos diferentes tipos de dados submetidos ao sistema. Nesse sentido, é necessário utilizar um padrão de representação que tenha capacidade para armazenar tipos distintos de informações. Alguns dos possíveis padrões e suas características são indicados na próxima seção.

### **4.3 Repositório de características**

Na camada denominada Repositório de características estão localizados todos os componentes relacionados ao armazenamento, visualização e acesso aos dados obtidos através do processamento das entradas. O elemento central dessa camada contém o padrão de representação a ser utilizado. Com o objetivo de suprir a segunda lacuna de pesquisa, que trata da interoperabilidade entre sistemas de saúde, é necessário empregar um padrão de representação reconhecido atualmente. Além disso, outro importante benefício é a possibilidade de armazenar as informações obtidas, através do processamento de diferentes tipos de dados de entrada, de maneira organizada e estruturada, permitindo assim a sua utilização por sistemas computacionais. Foram citados alguns desses padrões no Capítulo 2, com destaque para as especificações do openEHR e do HL7 FHIR.

Dados comuns a serem armazenados no repositório de características, independente do sistema a ser desenvolvido, estão relacionados aos pacientes e às observações referentes a eles. A Tabela 5 apresenta algumas dessas informações e os elementos correspondentes que podem ser utilizados para representá-las, tanto no openEHR como no HL7 FHIR. A partir das informações contidas na tabela citada, é possível verificar as diferentes formas disponíveis para se representar uma mesma informação em cada um dos padrões. Essa condição pode prejudicar a utilização do modelo, uma vez que exige que cada pesquisa a ser agregada no sistema seja desenvolvida de acordo com as especificações do padrão de representação

utilizado, aumentando assim a complexidade para incorporar pesquisas já desenvolvidas e, conseqüentemente, enfraquecendo a característica expansível do modelo proposto. Outro ponto negativo está relacionado a uma eventual substituição do padrão de representação. Se isso for preciso, seria necessário modificar todos os elementos que inserem dados no repositório (camada de extração de características) e todos os elementos que utilizam dados existentes no repositório (camada de aplicações), enfraquecendo a característica modular do modelo proposto.

**Tabela 5: Exemplos de dados e suas correspondências nos padrões openEHR e HL7 FHIR.**

Dado		HL7 FHIR	openEHR			
Classe	Atributo	Resource	Campo	Formato	Archetype	Item
Paciente	Nome	<i>Patient</i>	<i>name</i>	<i>1x Family</i>	<i>person_name.v1</i>	<i>1x given</i>
				<i>*x given</i>		<i>*x middle</i>
						<i>1x family</i>
	Nascimento		<i>birthDate</i>		<i>individual_personal.v1</i>	<i>Date of Birth</i>
	Gênero		<i>gender</i>			<i>Sex</i>
	Endereço		<i>address</i>	<i>*x line</i>	<i>address.v1</i>	<i>1..4x line</i>
				<i>1x city</i>		
				<i>1x state</i>		<i>1x number</i>
				<i>1x country</i>		
Observação	Tipo	<i>Observation</i>	<i>Comment</i>		<i>clinical_evidence.v1</i>	<i>Finding</i>
	Valor		<i>Value</i>			<i>Result</i>

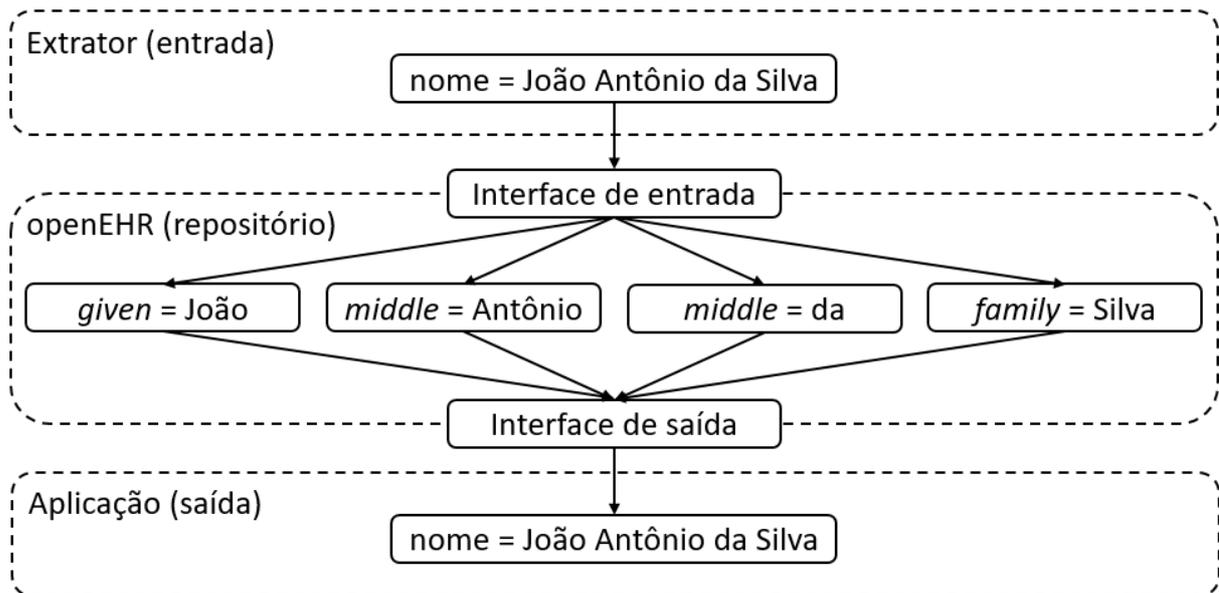
Fonte: Elaborado pelo autor.

Com o objetivo de isolar a complexidade relacionada ao tipo de repositório utilizado, removendo os fatores negativos citados, foram definidos os componentes denominados interfaces de entrada e de saída. Esses elementos contêm as ligações entre uma lista pública de informações disponíveis para uso, como as indicadas nas duas primeiras colunas da Tabela 5, com os elementos correspondentes de acordo com o padrão de representação selecionado para uso. Como exemplo pode ser citado o nome de um paciente. Na interface de entrada está disponível um método para inserir o nome de um paciente, e na interface de saída está disponível um método para retornar o nome de um paciente. A forma como essa informação será armazenada no repositório de características não é relevante para os demais elementos do modelo, sendo responsabilidade das interfaces de entrada e de saída adaptar o conteúdo ao padrão utilizado. Se for utilizado, por exemplo, o padrão openEHR, é a interface de entrada que deverá separar o nome em *given*, *middle* e *family*, para fique de acordo com a especificação. Da mesma forma, quando a aplicação solicitar essa informação, é papel da interface de saída unir os campos *given*, *middle* e *family*, transformando-os em um único dado. A Figura 29 ilustra esse cenário.

A existência de uma interface de entrada e de saída no modelo flexibiliza o emprego de outros métodos de armazenamento para os dados. Modificando as interfaces citadas, o modo de representação pode ser substituído sem que sejam necessárias alterações nos componentes dos módulos anteriores (extração de características) e nos componentes dos módulos posteriores (aplicações). Para isso, basta que seja possível relacionar as estruturas de dados existentes na base em uso atualmente com as estruturas de dados utilizadas na especificação que se deseja utilizar. Retomando o exemplo referente ao nome do paciente, se

o desejo for substituir o uso do padrão openEHR pelo padrão HL7 FHIR, não será necessário alterar todos os elementos que fazem uso dos dados existentes no repositório de características, já que é preciso modificar apenas os elos de comunicação entre as interfaces e o padrão de representação. Isso é possível uma vez que os demais elementos do modelo abordam a manipulação do nome como um dado único (essa condição também pode ser observada na Figura 29). Essa característica complementa o preenchimento da terceira lacuna citada, possibilitando a adaptação do modelo para diferentes padrões de armazenamento e representação.

**Figura 29: Simplificação obtida através do uso das interfaces de entrada e de saída.**



Fonte: Elaborado pelo autor.

De acordo com o padrão escolhido para ser empregado no repositório de características, pode ser necessário o domínio sobre a sua especificação para que seja possível visualizar as características disponíveis, além da forma em que elas estão organizadas e relacionadas. Com o objetivo de reduzir o conhecimento necessário sobre o padrão utilizado para este fim, é indicado o uso de ontologias como forma alternativa para acesso ao conteúdo do repositório. Esse tipo de representação tende a ser mais amigável e intuitivo para pessoas que não estão familiarizadas com os recursos e seus atributos correspondentes especificados no HL7 FHIR, por exemplo.

A Figura 30 apresenta, em XML, um fragmento de um laudo médico no padrão HL7 FHIR. Nesse exemplo pode-se perceber que foram utilizados os atributos “*id*”, cujo valor é “45059”, “*status*”, cujo valor é “*partial*” e “*issued*”, com valor igual a “2017-09-17T17:30:21.987-03:00”. Também constam os relacionamentos utilizados. O primeiro possui uma referência a um recurso “*patient*”, indicado pelo atributo “*subject*” que aponta para o paciente 35052. Os demais são referências às observações do laudo, com “*ids*” iguais a 45056, 45057 e 45058, respectivamente. Essa forma de representação dos dados pode ser confusa para pessoas que desconhecem os detalhes do padrão, além de ser necessário algum nível de conhecimento sobre a notação XML.

**Figura 30: Representação em XML de um fragmento de laudo no padrão HL7 FHIR.**

```
<DiagnosticReport xmlns="http://hl7.org/fhir">
  <id value="45059"/>
  <status value="partial"/>
  <subject>
    <reference value="Patient/35052"/>
  </subject>
  <issued value="2017-09-17T17:30:21.987-03:00"/>
  <result>
    <reference value="Observation/45056"/>
  </result>
  <result>
    <reference value="Observation/45057"/>
  </result>
  <result>
    <reference value="Observation/45058"/>
  </result>
</DiagnosticReport>
```

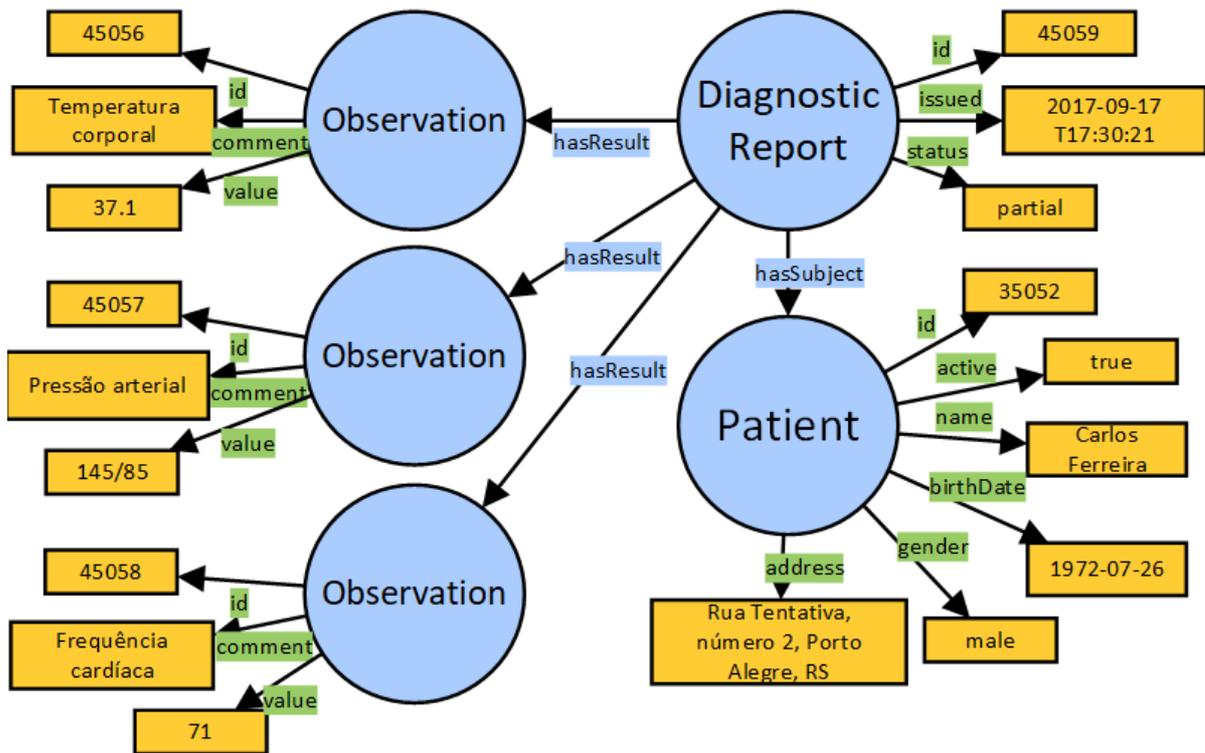
Fonte: Elaborado pelo autor.

A Figura 31, por sua vez, ilustra o mesmo laudo médico exibido na Figura 30, porém na forma de ontologia. Nesse caso, todas as informações são apresentadas de maneira gráfica. Os recursos são representados pelos círculos azuis, os relacionamentos pelas flechas com texto azul, os atributos têm seu nome nas flechas com texto em verde e seus respectivos valores estão nas caixas amarelas. Nessa forma de visualização são exibidas todas as informações referentes ao laudo que aparecem na representação textual (XML). A representação também contém os atributos de cada uma das classes referenciadas (paciente e observações, no exemplo), permitindo uma visão geral de todos os recursos utilizados.

A ontologia que retrata o repositório de características, se for utilizado o padrão HL7 FHIR, pode ser construída com base nas especificações já existentes na documentação dessa especificação, que descreve os recursos no padrão RDF<sup>22</sup>, ou ser fundamentada no estudo apresentado por Martinez-Costa e Schulz (2017). Acredita-se que, através da visualização dessa estrutura e dos dados correspondentes a cada um dos recursos, o desenvolvimento de funcionalidades para o sistema pode ser facilitado, uma vez que, como citado anteriormente, este artefato reduz o conhecimento necessário sobre a estrutura da base de dados utilizada.

<sup>22</sup> HL7 FHIR RDF Representation – <https://www.hl7.org/fhir/rdf.html>

Figura 31: Ontologia correspondente ao fragmento de laudo apresentado na Figura 30.



Fonte: Elaborado pelo autor.

O último elemento que compõe a camada do repositório de características é denominado Bases externas, representando uma porta de entrada e de saída das informações armazenadas no repositório. O objetivo é possibilitar a integração com bases de dados externas, viabilizando assim a disponibilização dos atributos extraídos para serem empregados em sistemas de terceiros. Também existe a viabilidade da utilização, por parte dos demais elementos do modelo, de informações disponíveis em bases de conhecimento que não estão, necessariamente, estabelecidas no mesmo local físico do repositório de características. Esse componente foi observado nos trabalhos de Puppala et al. (2015) e Mourão, Martins e Magalhães (2014), citados como relacionados.

#### 4.4 Aplicações

O último grupo de elementos existente no modelo contém as aplicações que fazem uso dos dados armazenados no repositório de características. Na Figura 28 foram indicados alguns exemplos de aplicações que podem utilizar as características do modelo, como os sistemas de apoio ao diagnóstico, geração de laudos, conferência de laudos, recomendação de medicamentos e indexação de exames. A seguir são comentados detalhes de algumas destas aplicações. Estes exemplos foram definidos a partir do estudo da literatura, de entrevistas com profissionais da área médica e também com base na avaliação do autor sobre possibilidades técnicas.

A aplicação denominada “Conferência de laudos” tem, como entrada, as imagens de determinado exame e o laudo correspondente, produzido pelo profissional responsável pela sua análise. As imagens são processadas, buscando encontrar as características relevantes. Em

paralelo a isso, o texto do laudo também passa por um algoritmo que identifica e extrai as informações importantes. Posteriormente, os dados obtidos através dos dois processos são comparados, possibilitando determinar se, por exemplo, todas as características extraídas das imagens estão sendo referenciadas no texto. Além da existência, caso sejam utilizados valores numéricos, é possível verificar se estão iguais aos encontrados no processamento das imagens. Qualquer diferença localizada pode gerar uma notificação ao especialista, indicando que determinado item deve ser revisado. Esse processo é especialmente importante pois, conforme citado por Kansagra et al. (2015), o profissional responsável pela análise das imagens e redação do laudo de um exame pode desconsiderar informações que possuem grande relevância para o médico solicitante, já que podem possuir especialidades e métodos de avaliação diferentes.

Outra operação de conferência a qual laudos médicos podem ser submetidos está relacionada à revisão do processo de transcrição de áudio para texto. A entrada, nesse caso, é formada pelo áudio gravado pelo médico ou especialista e o laudo correspondente, que foi digitado por outro integrante da equipe. No processamento, o áudio que contém as observações feitas pelo profissional em relação ao exame é transformado em texto a partir de um sistema de reconhecimento de voz. A aplicação contém também a etapa responsável pelo processamento do texto resultante, que tem como finalidade a extração das informações do laudo transcrito, permitindo assim a execução de uma etapa que verifica se as frases estão dentro do contexto do exame, evitando que sentenças e palavras impróprias, ou incorretas, sejam mantidas no texto.

Pode ser interessante, em casos relacionados ao ensino ou uso das informações por pessoas que não possuem o conhecimento pleno sobre o tema, a utilização de um sistema capaz de identificar as características citadas em um laudo médico e as destacar nos exames correspondentes. Esse tipo de aplicação tem, como entrada, as imagens que compõem o exame e o seu laudo. O texto é percorrido com o objetivo de extrair as observações que abordam determinada região, seja a descrição de uma lesão ou outra condição passível de ser visualizada. As características detectadas no texto são então utilizadas como entrada para o elemento responsável pelo processamento das imagens, que busca localizar a região correspondente e, posteriormente, destacar as respectivas características, facilitando a sua avaliação.

Com o objetivo de otimizar o processo de construção de laudos médicos pode ser utilizado um sistema que produz um laudo preliminar, a partir do processamento e identificação das características relevantes em exames por vídeo, imagem ou som, além dos dados de sensores. As informações importantes são inseridas, na forma textual, em modelos de laudos. Cada característica encontrada pode estar relacionada a uma frase pré-cadastrada, que é automaticamente determinada e inserida no texto. Ao final do processo, é possível que o laudo gerado contenha a maior parte das informações necessárias, de modo que o profissional responsável fica encarregado pela sua revisão, complementação e liberação.

Fora da linha de sistemas utilizados para auxiliar os profissionais da saúde, o modelo também pode ser a base para uma aplicação que contribui na tarefa de verificação da eficácia de novos métodos e algoritmos relacionados ao processamento de dados clínicos, de maneira semelhante ao que foi descrito por Szlosek e Ferrett (2016) em sua pesquisa. Métodos manuais, segundo os autores citados, exigem um grande volume de esforço braçal, não sendo assim eficientes. A extração de características de diferentes fontes de dados pode contribuir para a validação de sistemas de apoio, permitindo o processamento e comparação de grandes volumes de informações de maneira automatizada.

Todos os exemplos citados, que abordam possíveis tipos de aplicações a serem desenvolvidas com base no modelo proposto, apresentam as suas diversas formas de uso. A partir da extração de características existentes em tipos de dados heterogêneos e seu armazenamento em um repositório padronizado, diferentes sistemas podem obter e empregar as informações disponibilizadas com o objetivo de melhorar o atendimento prestado aos pacientes. A partir da descrição dessas aplicações, que possuem diferentes finalidades, mas que podem fazer uso de características já inseridas no repositório a partir de outras aplicações, é demonstrado o preenchimento da terceira lacuna de pesquisa identificada nos trabalhos relacionados, que trata da possibilidade de estender o emprego do sistema de maneira simplificada.

#### **4.5 Considerações**

O detalhamento do modelo apresentado não define as especificações técnicas relacionadas à implementação de cada um dos seus componentes, como por exemplo, os padrões, técnicas e métodos a serem empregados nas rotinas de extração de características, pois essas questões estão relacionadas diretamente ao tipo de entrada a ser utilizada, as informações que se deseja extrair e a finalidade das aplicações. Da mesma forma, não é especificado o padrão de representação a ser utilizado.

A estrutura do modelo apresentado foi formulada com foco na modularização e possibilidade de expansão dos sistemas desenvolvidos, além de permitir seu uso em diferentes tipos de aplicações. No capítulo seguinte estão descritos os dois protótipos desenvolvidos, que fazem uso da estrutura apresentada no modelo proposto, a partir do detalhamento das técnicas e dos padrões utilizados.

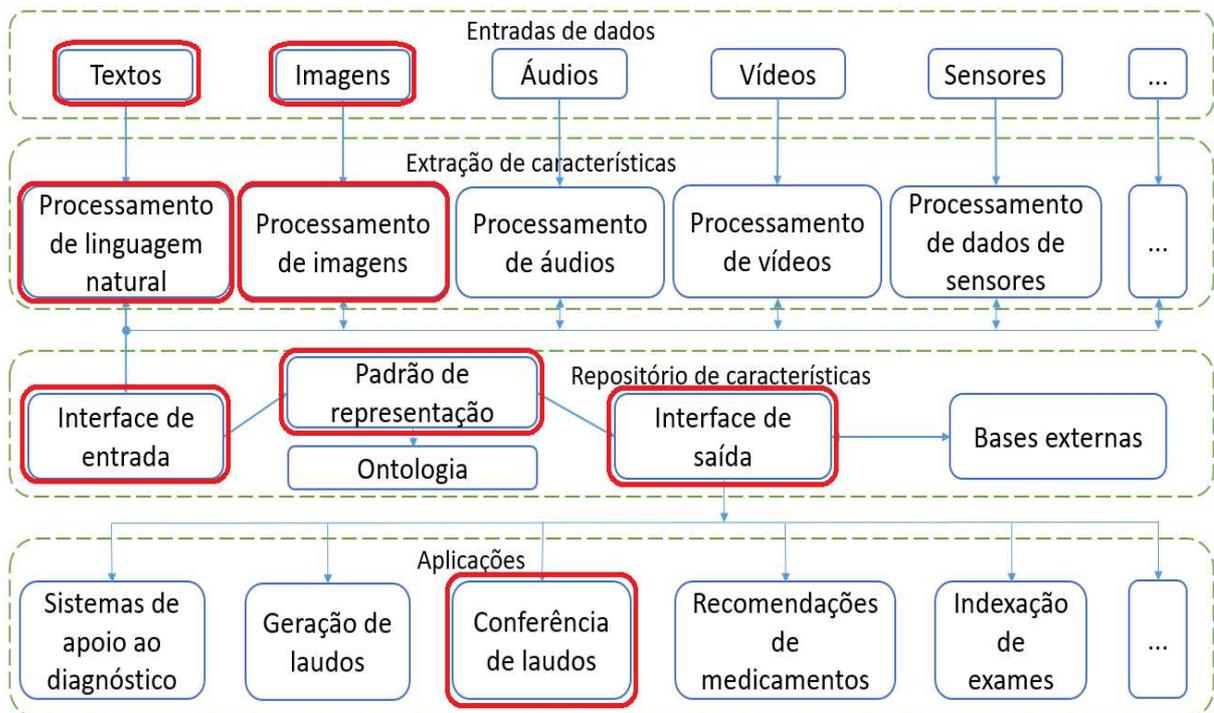
## 5 DESENVOLVIMENTO DOS PROTÓTIPOS

Com a finalidade de verificar a capacidade de utilização e integração dos itens que formam o modelo descrito no capítulo anterior foram desenvolvidos dois protótipos. Esses sistemas contêm os elementos fundamentais para exemplificar o uso e possibilitar a realização de provas de conceito do modelo apresentado. O primeiro deles consiste em uma aplicação que faz a conferência de laudos de exames de tomografia computadorizada do tórax. O segundo é uma aplicação que gera um laudo preliminar do estado de saúde de um paciente, com base em características extraídas de exames por imagem e dados provenientes de sensores.

### 5.1 Conferência de laudos

A aplicação desenvolvida neste protótipo permite detectar, em laudos médicos de exames por imagem, se informações importantes não foram citadas e se dados foram inseridos de maneira incorreta. Ao ser identificada uma ou mais dessas ocorrências, é gerado um alerta para o profissional responsável, indicando que determinado conteúdo do texto deve ser revisado. Esse tipo de aplicação pode contribuir para o aumento da qualidade dos laudos, reduzindo eventuais falhas e garantindo que contenham todas as informações relevantes.

**Figura 32: Elementos do modelo que foram desenvolvidos no primeiro protótipo.**



Fonte: Elaborado pelo autor.

A Figura 32 apresenta os componentes do modelo, destacados em vermelho, que formam o primeiro protótipo construído. Foram consideradas, como entradas, dados na forma de textos e dados na forma de imagens, envolvendo assim os módulos de Processamento de

linguagem natural e de Processamento de imagens. O desenvolvimento deste protótipo foi dividido em quatro etapas, cada uma correspondendo a um ou mais elementos que compõem o sistema. As etapas são detalhadas nas subseções a seguir.

### 5.1.1 Repositório de características

O primeiro componente criado no protótipo foi o repositório de características. A escolha do padrão a ser utilizado foi feita com base na pesquisa de Allwell-Brown (2016), que explorou e comparou os padrões HL7 FHIR e openEHR, possibilitando a identificação de similaridades, diferenças, vantagens e fraquezas correspondentes a cada um dos métodos de armazenamento e representação de dados clínicos. A partir da avaliação dos resultados obtidos pelo autor citado, que envolvem critérios como flexibilidade, simplicidade, escalabilidade, consistência na preservação semântica, formalismos de dados, legibilidade, reutilização de código, sintaxe de consulta, entre outras, o padrão escolhido para ser utilizado nos protótipos foi o HL7 FHIR, uma vez que este possui maior simplicidade, leveza e velocidade no aprendizado e implementação. Outras características, conforme Allwell-Brown (2016), são: maior formalidade e coerência na representação dos dados clínicos (tanto em XML como em JSON); boa legibilidade nas sentenças de consultas à base de dados, tanto as simples como as complexas, a partir do uso da interface REST; mecanismos mais robustos e completos para emprego de terminologias externas.

A implementação do padrão de representação, feita de acordo com a especificação HL7 FHIR, se divide em duas aplicações: cliente (denominado FHIRClient) e servidor (denominado FHIRJPAServer). Em ambas foi utilizada a biblioteca HAPI FHIR<sup>23</sup>, que é uma implementação em Java da especificação HL7 FHIR, de código aberto e uso livre. O FHIRJPAServer é responsável por gerenciar o armazenamento das informações e por responder às operações de inserção, alteração, consulta e deleção dos dados. Esse item foi construído com base no exemplo *JPA Server*<sup>24</sup>, disponibilizado pelos autores da biblioteca HAPI FHIR, que faz uso banco de dados *Apache Derby*<sup>25</sup> e é executado no servidor *Apache Tomcat*<sup>26</sup>. Foram feitas alterações relacionadas ao endereço de acesso e denominação do servidor. Esse componente possui um portal *web* que viabiliza acesso direto aos dados. A Figura 33 apresenta a página inicial desse portal, que pode ser acessada a partir de qualquer navegador, de maneira local, pelo endereço <http://localhost:8084/FHIRJPAServer>.

O menu localizado na lateral esquerda da página, apresentada na Figura 33, viabiliza o acesso às instâncias de todos os recursos disponíveis no HL7 FHIR e que foram inseridos na base, tais como pacientes, apontamentos, condições, dispositivos, entre outros. Na área específica de cada recurso é possível executar pesquisas ou editar os dados correspondentes. Entretanto, esse método é utilizado somente para testes, uma vez que o objetivo do modelo é permitir que outras aplicações possam utilizar as informações disponíveis no repositório de forma direta e automatizada.

---

<sup>23</sup> HAPI FHIR – <http://hapifhir.io/>

<sup>24</sup> JPA Server – <https://github.com/jamesagnew/hapi-fhir/tree/master/hapi-fhir-jpaserver-example>

<sup>25</sup> Apache Derby – <https://db.apache.org/derby/>

<sup>26</sup> Apache Tomcat – <http://tomcat.apache.org/>

**Figura 33: Exemplo de página do portal de acesso ao repositório de dados.**

The screenshot shows the HAPI-FHIR web portal. The header includes 'Home', 'Server: Local Tester', 'Source Code', and 'About This Server'. The left sidebar contains 'Options' (Encoding: default, XML, JSON; Pretty: default, On, Off; Summary: none, true, text, data, count), 'Server' (Server Home/Actions), and 'Resources' (Patient, Account, ActivityDefinition, AdverseEvent, AllergyIntolerance, Appointment). The main content area features the HAPI-FHIR logo and the text 'HAPI-FHIR fhir made simple.' Below this is a warning: 'This is not a production server! Do not store any information here that contains personal health information or any other confidential information. This server will be regularly purged and reloaded with fixed test data.' A table provides server details:

Server	Servidor Dissertação
Software	HAPI FHIR Server - 2.5-SNAPSHOT
FHIR Base	http://localhost:8084/FHIRJPAServer/baseDstu3

Below the table is a section for 'Server Actions'.

Fonte: Elaborado pelo autor.

A aplicação que possibilita o acesso direto às informações existentes na base de dados a partir dos demais itens do modelo é o FHIRClient, viabilizando assim a execução de operações de inserção, alteração e consulta. Esse componente corresponde aos elementos denominados Interface de entrada e Interface de saída, que foram descritos no modelo. O projeto utilizado como base para a construção desse componente foi o *RESTful Generic Client*<sup>27</sup>, disponibilizado pelos autores da biblioteca HAPI FHIR. Com o objetivo de servir como um intermediário entre as aplicações e o repositório de características, foram adicionados ao FHIRClient todos os métodos necessários para conexão com o FHIRJPAServer e para manipulação das informações. A Tabela 6 contém a lista das funções desenvolvidas e uma breve descrição sobre o emprego de cada uma.

**Tabela 6: Métodos disponíveis para uso no FHIRClient.**

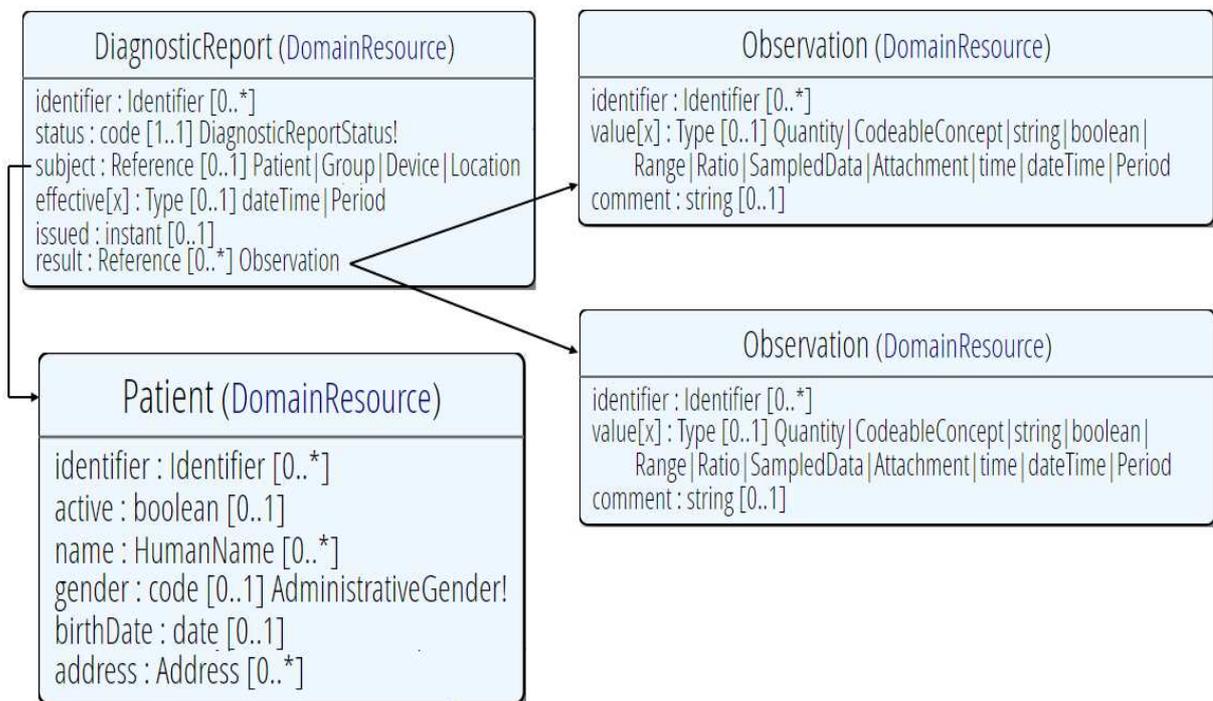
Método	Descrição
insertPatient	Utilizado na inserção do recurso <i>Patient</i> no repositório de características
readPatient	Utilizado na leitura do recurso <i>Patient</i> do repositório de características
insertDiagnosticReport	Utilizado na inserção do recurso <i>DiagnosticReport</i> no repositório de características
readDiagnosticReport	Utilizado na leitura do recurso <i>DiagnosticReport</i> do repositório de características
readPatientDR	Utilizado para extração da referência correspondente ao recurso <i>Patient</i> utilizado no recurso <i>DiagnosticReport</i>
readCalcioPDI	Utilizado para leitura do valor correspondente ao recurso <i>Observation</i> utilizado no recurso <i>DiagnosticReport</i> , que representa a existência de calcificações nas imagens
readCalcioPLN	Utilizado para leitura do valor correspondente ao recurso <i>Observation</i> utilizado no recurso <i>DiagnosticReport</i> , que representa a descrição de calcificações no laudo

Fonte: Elaborado pelo autor.

<sup>27</sup> *RESTful Generic Client* – [http://hapifhir.io/doc\\_rest\\_client.html/](http://hapifhir.io/doc_rest_client.html/)

O principal recurso disponibilizado no HL7 FHIR que é utilizado neste protótipo é o *DiagnosticReport*, que contém as informações referentes a um exame. Como cada exame está relacionado a um paciente específico, foi utilizado também o recurso *Patient*. Cada instância de *DiagnosticReport* possui, no elemento *Subject*, a referência ao *Patient* correspondente. Além disso, outro recurso que foi utilizado é o *Observation*, que armazena dados específicos sobre os achados do exame. A característica avaliada no protótipo é a existência de pontos de calcificação na região cardíaca. Para representar essa informação são utilizadas duas instâncias de *Observation*. Na primeira é indicado se foram encontrados pontos de calcificação no processamento das imagens. A segunda é utilizada para indicar se foi encontrada referência à essa condição no texto do laudo. As etapas descritas nos itens 5.1.2 e 5.1.3 detalham a obtenção dessas informações. A Figura 34 ilustra os recursos do padrão HL7 FHIR utilizados, seus atributos e relacionamentos.

**Figura 34: Recursos, atributos e relacionamentos do HL7 FHIR utilizados.**



Fonte: Elaborado pelo autor.

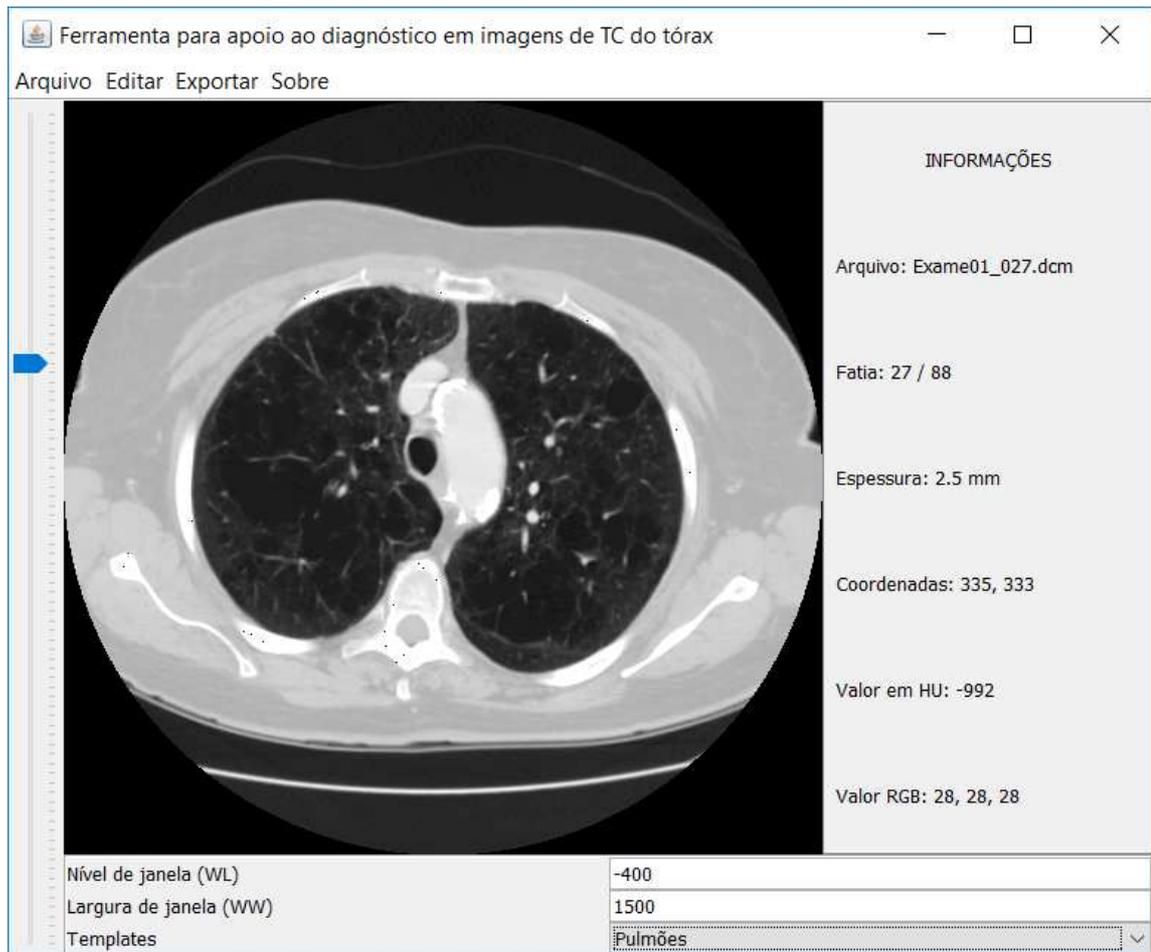
### 5.1.2 Processamento de imagens médicas

A segunda etapa de construção do protótipo envolveu a implementação relacionada a entrada e processamento de imagens médicas que compõem um exame de tomografia computadorizada do tórax. Para possibilitar o acesso a arquivos no padrão DICOM foi empregada a biblioteca *PixelMed Java DICOM Toolkit*<sup>28</sup>, também de código aberto e uso livre. Já o processamento das imagens tem como base a ferramenta descrita por Reis et al. (2016), que contém recursos para seleção e visualização de exames, navegação entre as imagens, alterações de nível e largura de janela, identificação e destacamento de estruturas

<sup>28</sup> *PixelMed Java DICOM Toolkit* – <http://www.pixelmed.com/dicomtoolkit.html>

anatômicas, entre outras funcionalidades. Além das opções já existentes, foi incluído nesse sistema um menu adicional, denominado “Exportar”. Ao acionar essa opção é disparado o processo de extração de características do exame em uso na ferramenta e, ao final do processo, são utilizados os métodos disponíveis no FHIRClient para inserir os dados obtidos no repositório de características. A Figura 35 ilustra a interface de uso da ferramenta de extração de características em exames de TC do tórax.

**Figura 35: Ferramenta para extração de características em exames de TC do tórax.**



Fonte: Elaborado pelo autor.

Uma vez que há um número elevado de condições que podem ser determinadas a partir da análise de exames de tomografia computadorizada do tórax, foi definido que apenas uma característica seria selecionada para a validação do modelo. A condição escolhida é descrita por Flores (2015), que aborda a existência de pontos de calcificação arterial coronariana. Os passos para detecção desses pontos envolvem a delimitação da região do pericárdio, cujo algoritmo foi detalhado por Kochenborger (2013), a aplicação de limiar para manter somente as áreas candidatas e a rotulação das regiões resultantes. Após a delimitação de cada segmento que pode representar cálcio, são descartados todos que não possuem uma área mínima, eliminando assim aqueles que não são relevantes ou que representam apenas ruído. Ao final da detecção, a existência ou não de áreas com cálcio é determinada e inserida no repositório de características, na primeira instância do recurso *Observation* vinculado ao *DiagnosticReport*.

### 5.1.3 Processamento de Linguagem Natural

A etapa seguinte consiste no processamento do conteúdo do laudo correspondente às imagens avaliadas na etapa anterior. Esta fase tem como finalidade identificar trechos do texto em linguagem natural que indiquem menções à presença de calcificações. O fluxo de ações descritas nesta etapa do processo é citado por Cai et al. (2016).

Inicialmente, o segmento do texto que descreve a interpretação do exame é submetido ao analisador sintático PALAVRAS (BICK, 2000), que separa as sentenças e disponibiliza as palavras normalizadas e classificadas em seus aspectos sintáticos e morfológicos. A Figura 36 apresenta um trecho do XML retornado pelo *parser* PALAVRAS.

**Figura 36: Trecho do XML retornado pelo *parser* PALAVRAS.**

Frase: Aorta de calibre normal, com placas de ateroma calcificadas.

```
<treebank id="Arboretum">
<body>

  <word id="1" form="Aorta" base="aorta" postag="n" morf="--" extra="* an tube" head="0" deprel="X"/>
  <word id="2" form="de" base="de" postag="prp" morf="--" extra="np-close" head="1" deprel="DN"/>
  <word id="3" form="calibre" base="calibre" postag="n" morf="--" sem="f-q" extra="--" head="2" deprel="DP"/>
  <word id="4" form="normal" base="normal" postag="adj" morf="--" extra="nh np-close" head="3" deprel="DN"/>
  <word id="5" form="," base="--" postag="pu" morf="--" extra="--" head="0" deprel="PU"/>
  <word id="6" form="com" base="com" postag="prp" morf="--" extra="np-close" head="3" deprel="Dnc"/>
  <word id="7" form="placas" base="placa" postag="n" morf="--" sem="cc-r cc-board" extra="--" head="6" deprel="DP"/>
  <word id="8" form="de" base="de" postag="prp" morf="--" extra="np-close" head="7" deprel="DN"/>
  <word id="9" form="ateroma" base="ateroma" postag="n" morf="--" sem="sick" extra="--" head="8" deprel="DP"/>
  <word id="10" form="calcificadas" base="calcificar" postag="v-fin" morf="--" extra="vN np-long" head="7" deprel="DN"/>
  <word id="11" form="." base="--" postag="pu" morf="--" extra="--" head="0" deprel="PU"/>

</sentence>
</body>
</treebank>
```

Fonte: Elaborado pelo autor.

Todas as frases que não contém o termo “calci” são desconsideradas. Nas frases restantes, os substantivos são verificados, buscando validar se estão relacionados às regiões anatômicas que devem ser consideradas. Essa ação é feita através do léxico *UMLS Metathesaurus Browser*<sup>29</sup>, onde são retornados termos equivalentes ao avaliado, reduzindo assim a lista de regiões anatômicas de interesse que a aplicação precisa manter. A última etapa consiste em verificar se a informação não está sendo negada. Para isso, são executadas buscas por termos que indicam essa condição nas frases restantes, como, por exemplo, “sem”, “não”, “ausência”, “ausente”, entre outros. Se o trecho não estiver sendo negado, a referência à presença de calcificação é considerada válida e essa informação é inserida no repositório de características, na segunda instância do recurso *Observation* vinculado ao *DiagnosticReport*.

<sup>29</sup> *UMLS Metathesaurus Browser* – <https://uts.nlm.nih.gov/metathesaurus.html>

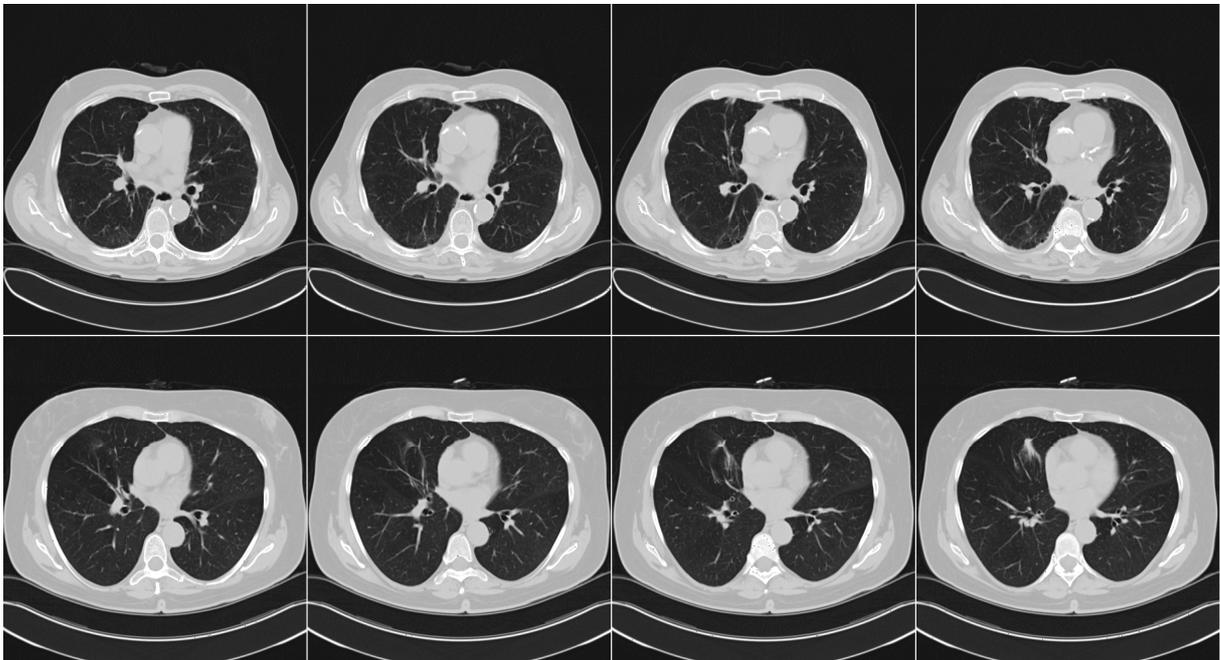
#### 5.1.4 Aplicação “Conferência de laudos”

O último elemento desenvolvido é a aplicação de conferência de laudos. A partir do identificador do exame que se deseja avaliar são feitas consultas ao repositório das características extraídas, possibilitando a comparação entre o resultado obtido pelo processamento das imagens e o resultado obtido pelo processamento do laudo. Essas consultas são viabilizadas através dos métodos disponíveis no FHIRClient, de modo que a aplicação não necessita conhecer detalhes da base de dados utilizada. Se forem detectadas diferenças nas observações geradas pelo processamento de cada tipo de entrada é exibido um alerta para o usuário indicando a situação identificada.

#### 5.1.5 Coleta de dados

Os dados utilizados na validação do protótipo estão relacionados aos dois tipos de entradas utilizadas: exames de tomografia computadorizada do tórax e os laudos médicos correspondentes. Os exames foram obtidos a partir da coleção *LungCT-Diagnosis*<sup>30</sup>, que é composta por tomografias computadorizadas do tórax pertencentes a 61 pacientes. Os arquivos DICOM estão disponíveis publicamente no repositório do TCIA (*The Cancer Imaging Archive*). A Figura 37 ilustra exemplos de imagens que compõem os exames obtidos na coleção *LungCT-Diagnosis*.

**Figura 37: Imagens de exames contidos na coleção *LungCT-Diagnosis*.**



Fonte: Elaborado pelo autor.

---

<sup>30</sup> *LungCT-Diagnosis* – <https://wiki.cancerimagingarchive.net/display/Public/LungCT-Diagnosis>

**Tabela 7: Laudos criados para testes de PLN.**

<b>Laudos</b>	<b>Informações clínicas</b>	<b>Interpretação</b>
1	Dor torácica esquerda. Tabagismo.	Enfisema centrolobular e parasseptal, predominando nos lobos superiores. Espessamento das paredes brônquicas. Ausência de derrame pleural. Linfonodos mediastinais menores que 1 cm no eixo transversal. Aorta de calibre normal, com placas de ateroma calcificadas. Coração de dimensões normais. Traqueia de calibre normal.
2	Dispneia.	Assimetria dos campos pulmonares, tendo dimensões menores à esquerda. Ausência de derrame pleural. Linfonodos em cadeias paratraqueais menores que 1,0 cm no eixo transversal. Coração de dimensões normais. Aorta de calibre normal. Traqueia de calibre preservado.
3	Tabagista. Uso de O <sub>2</sub> . Internado.	Enfisema centrilobular e parasseptal difusamente nos campos pulmonares. Espessamento difuso das paredes brônquicas. Espessamento pleural no ápice pulmonar esquerdo. Ausência de derrame pleural. Aorta de calibre normal, com placas de ateroma calcificadas. Coração de dimensões normais. Aorta de calibre normal.
4	Dispneia. Leucemia.	Moderado derrame pleural bilateral. Linfonodo em cadeia paratraqueal direita medindo 1,7 x 1,1 cm. Aorta de calibre preservado, ateromatosa. Coração de dimensões normais, com calcificações valvares. Traqueia pérvia, de calibre normal.
5	Controle de linfoma.	Persiste aumento da densidade da gordura mediastinal anterosuperior medindo cerca de 2,5 x 1,5 cm. Alterações fibroatelectásicas no lobo superior. Ausência de derrame pleural. Não observamos linfonodomegalias nas cadeias paratraqueais, pré-vascular e subcarinal. Coração de dimensões normais. Aorta de calibre normal. Traqueia pérvia, de calibre preservado.

Fonte: Elaborado pelo autor.

No TCIA não estão disponíveis os laudos correspondentes aos exames. Por essa razão foram utilizados laudos fictícios, gerados a partir de exemplos de documentos dessa natureza.

Como o objetivo da validação é comparar as informações extraídas do texto em relação às obtidas através do processamento das imagens, os laudos foram construídos com o intuito de abranger todas as combinações possíveis de resultados, retratando assim diferentes cenários de testes. A Tabela 7 contém os laudos criados, que serão utilizados como base para as rotinas de processamento de linguagem natural.

### 5.1.6 Cenário de uso

A abordagem utilizada na validação é baseada em cenários de uso, que descrevem situações em que a aplicação construída de acordo com o modelo pode ser empregada. A partir das respostas fornecidas pelo protótipo em relação ao detalhamento do cenário é possível verificar se o modelo atende às necessidades, além de viabilizar a identificação das suas contribuições e eventuais fragilidades.

Descrição do cenário:

Dois pacientes chegaram ao hospital para execução de um exame de tomografia computadorizada do tórax, conforme solicitação de seus respectivos médicos. O primeiro paciente, chamado João da Silva, mora na Rua de Teste, número 48, em Novo Hamburgo e nasceu em 06/06/1981. Já a segunda paciente se chama Maria das Flores, reside na Rua Experimento, número 84, em São Leopoldo e nasceu em 30/08/1988.

**Figura 38: Retorno no formato XML da consulta ao recurso *Patient* que representa a Maria.**

```
<Patient xmlns="http://hl7.org/fhir">
  <id value="30055"/>
  <meta>
    <versionId value="1"/>
    <lastUpdated value="2017-05-21T21:00:51.485-03:00"/>
  </meta>
  <text>
  <active value="true"/>
  <name>
    <family value="Flores"/>
    <given value="Maria"/>
    <given value="das"/>
  </name>
  <gender value="female"/>
  <birthDate value="1988-08-30"/>
  <address>
    <use value="home"/>
    <line value="Rua Experimento, numero 84"/>
    <city value="São Leopoldo"/>
    <state value="RS"/>
  </address>
</Patient>
```

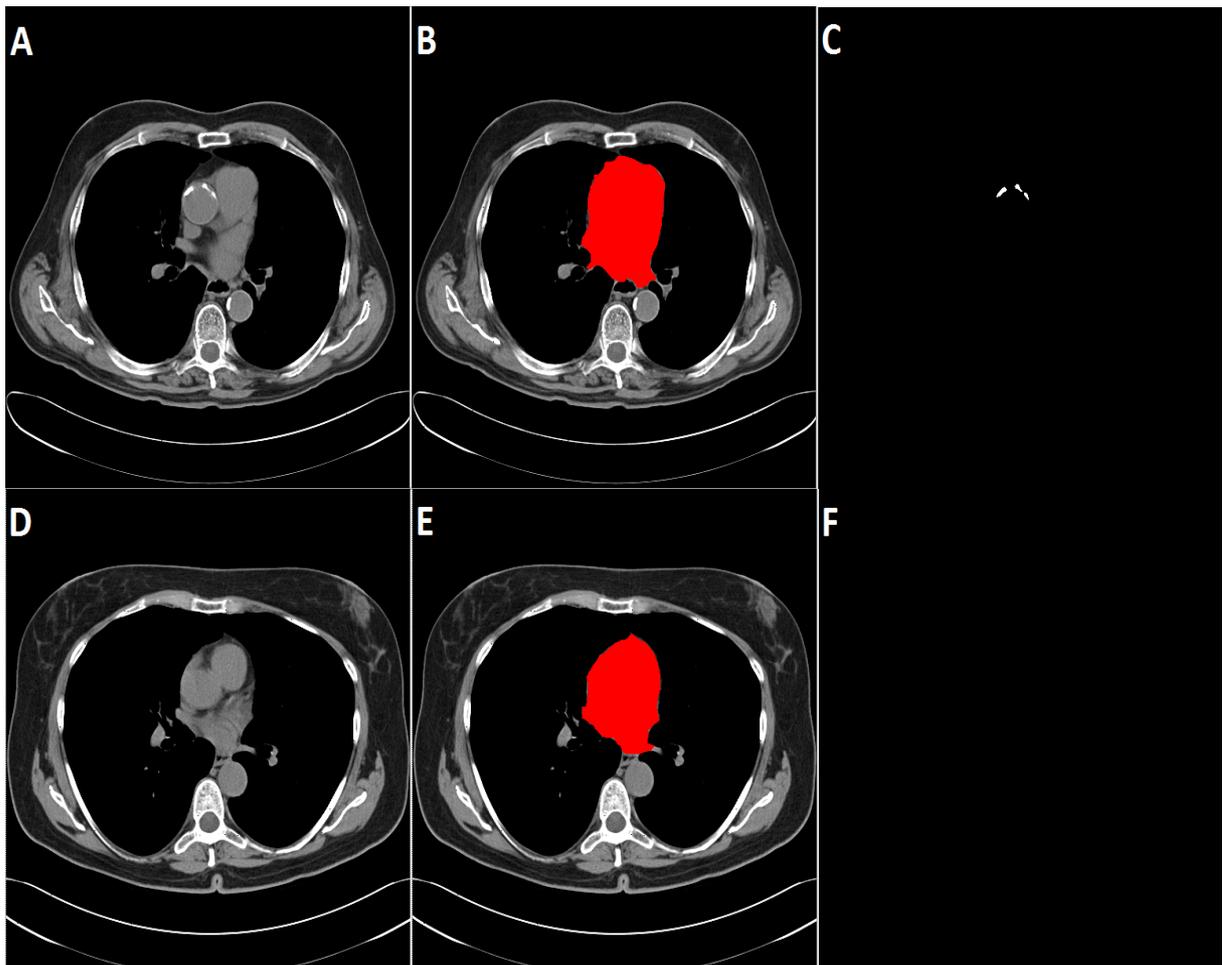
Fonte: Elaborado pelo autor.

A primeira ação necessária é a inclusão dos pacientes na base de dados. Para representá-los são utilizadas duas instâncias do recurso *Patient*. A inserção é feita através do componente FHIRClient, e a confirmação pelo portal *web* disponibilizado pelo FHIRJPAServer. O resultado da inclusão da paciente Maria é exibido na Figura 38. O identificador do recurso, gerado automaticamente, é indicado pela *tag* “*id*”, cujo valor é “30055”.

No campo “*text*” ficam as informações do recurso formatadas de maneira amigável para exibição (esse conteúdo foi omitido na Figura para possibilitar a exibição completa do restante do arquivo). Os outros dados indicam a situação do cadastro (*tag* “*active*”), o nome (conjunto de campos da *tag* “*name*”), o gênero (*tag* “*gender*”), a data de nascimento (*tag* “*birthDate*”) e o endereço (conjunto de campos da *tag* “*address*”). Estão disponíveis diversos outros atributos para serem associados a esse recurso, que podem ser utilizados se necessário. O identificador gerado para o paciente João foi “30054”.

Após o cadastro, os pacientes são encaminhados para a execução do exame. O tomógrafo gera um conjunto de imagens, no padrão DICOM, que é repassado para o profissional responsável pela análise e redação do laudo. Ao concluir a escrita, o texto e as imagens são submetidas ao protótipo, onde são executados os processos de extração de características correspondentes à cada tipo de entrada.

**Figura 39: Detecção de pontos de calcificação em imagens de TC.**



Fonte: Elaborado pelo autor.

A Figura 39 apresenta as etapas e os resultados do processo de extração de características das imagens, em que pontos de calcificação na região do pericárdio são detectados. A primeira linha (imagens A, B e C) é referente ao paciente João. A segunda linha (imagens D, E e F) pertence ao exame da paciente Maria. A primeira coluna (imagens A e D) apresenta as imagens originais. A segunda coluna (imagens B e E) destaca a região de interesse, em vermelho. A última coluna (imagens C e F) mostra somente os pontos de calcificação encontrados, na cor branca, dentro da região de interesse. É possível visualizar que existem pontos brancos, que representam calcificações, somente na imagem C, que pertence ao exame do paciente João.

Os relatórios correspondentes a cada exame ficaram bastante semelhantes na área denominada “Interpretação”, que contém os achados do responsável pela análise das imagens. A Figura 40 ilustra o laudo do paciente João. No caso da paciente Maria, o resultado também foi normal, sem que tenham sido descritas alterações. Como não existe nenhuma ocorrência do termo “calci” no laudo, o processo de extração de características não detectou referências em relação a existência de pontos de calcificação.

**Figura 40: Laudo resultante do paciente João.**

**Informações clínicas:**

Dor torácica esquerda; Tabagismo;

**Técnica:**

Exame realizado em aparelho helicoidal, sem a infusão do meio de contraste.

**Interpretação:**

Ausência de derrame pleural.

Não observamos linfonomegalias nas cadeias paratraqueais, pré-vascular e subcarinal.

Coração de dimensões normais.

Aorta de calibre normal.

Traquéia pérvia, de calibre preservado.

Estrutura óssea normal.

Fonte: Elaborado pelo autor.

Após o processamento das imagens e dos laudos correspondentes, são inseridos os registros relacionados a cada exame no repositório de características. Para isso são utilizadas duas instâncias do recurso *DiagnosticReport*, um para cada exame. No *DiagnosticReport* do paciente João, a primeira instância do recurso *Observation* conterà a informação de que pontos de calcificação foram identificados no processamento das imagens. No recurso correspondente utilizado para a paciente Maria não haverá esse dado. Já a segunda instância de *Observation*, para ambos os pacientes, conterà a informação de que nenhuma referência à existência de calcificações foi encontrada nos laudos. A Figura 41 exhibe o resultado, no formato XML, do *DiagnosticReport* inserido para representar o exame do paciente João. O identificador gerado para o recurso foi “30056”. Também foi omitido o conteúdo da tag “*text*”, pois ela contém as mesmas informações que o restante das tags, apenas formatada de maneira amigável para visualização.

É possível verificar que o *DiagnosticReport* está vinculado ao paciente João através do conteúdo da tag “*subject*”, que contém uma referência para o recurso “Patient/30054”, que é o identificador desse paciente, conforme citado anteriormente. Além disso, podem ser visualizadas as duas instâncias do recurso *Observation*, que contém, respectivamente, os valores de cálcio extraídos das imagens e do texto do laudo. Nesse caso, o valor 1 indica que existe referência à pontos de calcificações e o valor 0 nega a existência.

**Figura 41:** XML retornado na consulta ao recurso *DiagnosticReport* do paciente João.

```
<DiagnosticReport xmlns="http://hl7.org/fhir">
  <id value="30056"/>
  <meta>
    <versionId value="1"/>
    <lastUpdated value="2017-05-21T22:35:57.184-03:00"/>
  </meta>
  <text>
  <contained>
    <Observation xmlns="http://hl7.org/fhir">
      <id value="1"/>
      <valueQuantity>
        <value value="1"/>
      </valueQuantity>
      <comment value="Valor de cálcio extraído via PDI"/>
    </Observation>
  </contained>
  <contained>
    <Observation xmlns="http://hl7.org/fhir">
      <id value="2"/>
      <valueQuantity>
        <value value="0"/>
      </valueQuantity>
      <comment value="Valor de cálcio extraído via PLN"/>
    </Observation>
  </contained>
  <status value="partial"/>
  <subject>
    <reference value="Patient/30054"/>
  </subject>
  <effectiveDateTime value="2017-05-21T22:35:56-03:00"/>
  <issued value="2017-05-21T22:35:56.010-03:00"/>
  <result>
    <reference value="#1"/>
  </result>
  <result>
    <reference value="#2"/>
  </result>
</DiagnosticReport>
```

Fonte: Elaborado pelo autor.

Foi implantada mais uma etapa no processo do hospital, que trata da conferência dos laudos. Essa tarefa é executada por uma aplicação, de forma automatizada, que faz uso das

características extraídas e armazenadas no *DiagnosticReport*. Através do identificador do exame são recuperados e comparados os dados das instâncias de *Observation* existentes. Se os valores estiverem iguais, o laudo está correto e pode ser entregue para o paciente ou médico solicitante. Se estiverem diferentes é gerado um alerta, uma vez que há a chance de que alguma parte do processo tenha falhado e é aconselhável que o exame seja revisado. A Tabela 8 contém todas as possíveis combinações de valores que podem ser retornados e o resultado correspondente à verificação.

**Tabela 8: Possíveis valores e resultados da etapa de conferência dos laudos.**

Cálcio extraído via PDI	Cálcio extraído via PLN	Mensagem resultante
0	0	OK - Não foram citados pontos de calcificação no laudo e não foram identificados pontos de calcificação nas imagens.
0	1	REVISAR - Foram citados pontos de calcificação no laudo e não foram identificados pontos de calcificação nas imagens.
1	0	REVISAR - Não foram citados pontos de calcificação no laudo e foram identificados pontos de calcificação nas imagens.
1	1	OK - Foram citados pontos de calcificação no laudo e foram identificados pontos de calcificação nas imagens.

Fonte: Elaborado pelo autor.

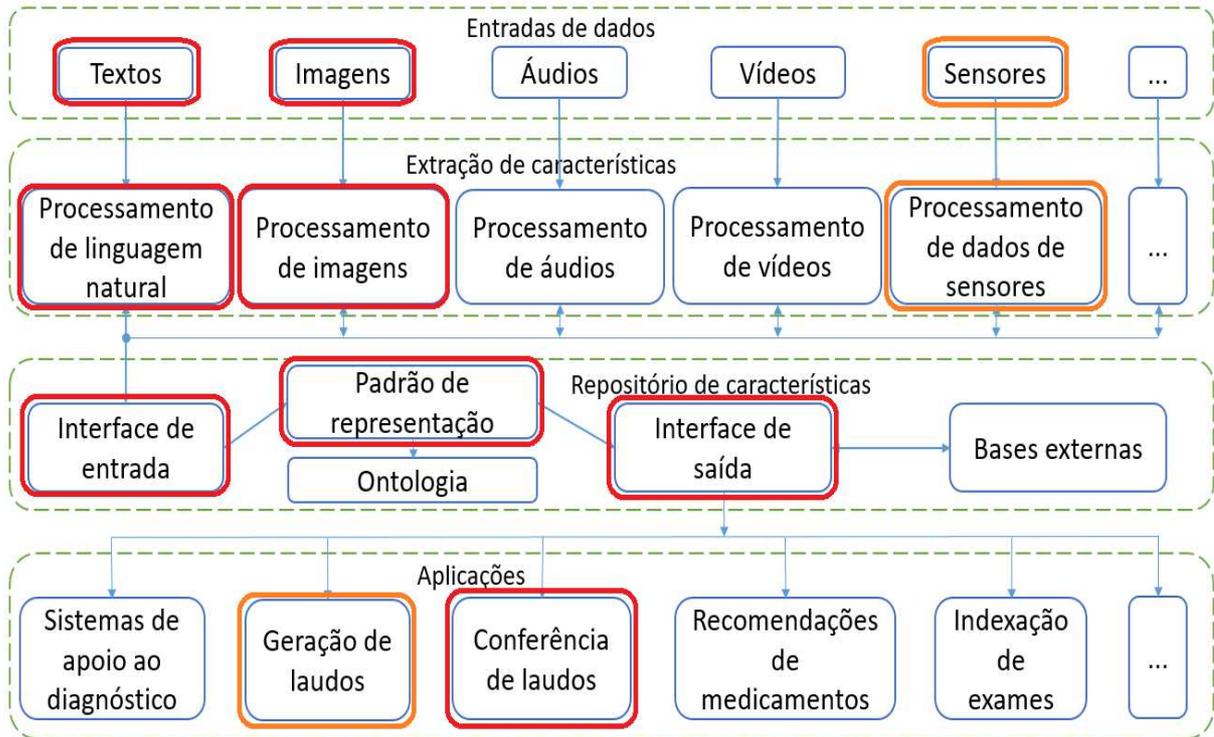
Os dois exames executados apresentaram resultados diferentes na etapa de conferência dos laudos. O exame do paciente João não contém informações sobre calcificações no laudo, mas foram detectados pontos na imagem. Sendo assim, foi exibido um alerta indicando essa condição e sugerindo a revisão do exame. Já no caso da paciente Maria, tanto o laudo como as imagens não demonstram essa condição, então a conferência indicou que o exame está correto e pode ser entregue.

## 5.2 Geração de laudo preliminar

O segundo protótipo desenvolvido consiste em uma aplicação que gera um laudo preliminar de determinado paciente, contendo frases referentes ao seu estado de saúde, a partir das informações obtidas através de exames por imagem e de sensores diversos. Esse tipo de aplicação pode contribuir para a otimização do processo de construção de laudos médicos.

A Figura 42 apresenta os componentes do modelo, destacados em laranja, que foram adicionados ao primeiro protótipo (elementos destacados em vermelho) e que formam o segundo protótipo construído. Foram adicionados, ao conjunto de entradas já existentes, dados provenientes de sensores de temperatura corporal, pressão sanguínea e frequência cardíaca.

**Figura 42: Elementos adicionados ao sistema na construção do segundo protótipo.**



Fonte: Elaborado pelo autor.

### 5.2.1 Repositório de características

Toda a estrutura correspondente ao repositório de características que foi desenvolvida no primeiro protótipo foi reaproveitada, corroborando a existência das facilidades que o modelo previa para estes casos. A partir da inclusão de um tipo adicional de entrada, composto por dados provenientes de sensores, foi necessário apenas adicionar os métodos de uso dessas informações no FHIRClient, que é o elemento correspondente às interfaces de entrada e de saída. A Tabela 9 descreve estes métodos adicionados.

**Tabela 9: Métodos adicionados ao FHIRClient para o segundo protótipo.**

Método	Descrição
readDRs	Utilizado na leitura de todos os <i>DiagnosticReports</i> de determinado paciente
readObservationTemp	Utilizado na leitura de uma instância do recurso <i>Observation</i> com o objetivo de retornar uma medida de temperatura
readObservationPress	Utilizado na leitura de uma instância do recurso <i>Observation</i> com o objetivo de retornar uma medida de pressão sanguínea
readObservationFreq	Utilizado na leitura de uma instância do recurso <i>Observation</i> com o objetivo de retornar uma medida de frequência cardíaca

Fonte: Elaborado pelo autor.

O recurso utilizado para armazenar os dados dos sensores é o mesmo que o utilizado no primeiro protótipo. Para o segundo protótipo, uma instância do *DiagnosticReport* pode

conter até 4 vinculações com instâncias do recurso *Observation*, sendo uma para a presença de calcificação definida via processamento de imagens, uma para a temperatura corporal, uma para a pressão sanguínea e uma para a frequência cardíaca. Também foi incluído no FHIRClient um método para retornar todos os *DiagnosticReports* de um mesmo paciente, já que o laudo preliminar pode ser composto por mais de um exame.

### 5.2.2 Processamento de dados de sensores

A etapa seguinte consiste no processamento dos dados provenientes de sensores. No protótipo não foram feitas integrações diretas com os dispositivos responsáveis por cada tipo de aferição, entretanto, em um sistema real essa é uma possibilidade que pode ser alcançada. Dessa forma, a rotina que processa as informações referentes à temperatura corporal, pressão sanguínea e frequência cardíaca é responsável por selecionar, de maneira aleatória em uma lista pré-definida de valores possíveis, as medidas a serem utilizadas.

### 5.2.3 Aplicação “Geração de laudo preliminar”

O último elemento desenvolvido é a aplicação de geração do laudo preliminar. A partir do identificador do paciente, todos os exames correspondentes ao código desejado são lidos. Para cada *DiagnosticReport* retornado, é feita a leitura dos recursos *Observation* que estão vinculados ao laudo lido. Todas as consultas feitas ao repositório de características são viabilizadas através dos métodos disponíveis no FHIRClient, de modo que a aplicação não necessita conhecer detalhes da base de dados utilizada.

**Tabela 10: Características disponíveis, regras e frases correspondentes.**

Característica	Regra	Frase inserida no laudo preliminar
Calcificação	Se = 0	Não foram detectados pontos de calcificação na região cardíaca.
	Se = 1	Foram detectados pontos de calcificação na região cardíaca.
Temperatura corporal	Se <= 37.5	Temperatura normal (VALOR °C).
	Se > 37.5	Febre (VALOR °C).
Frequência cardíaca	-	Frequência cardíaca: VALOR bpm.
Pressão sanguínea	S > 180 OU D > 110	Crise hipertensiva (VALOR mmHg).
	S >= 160 OU D >= 100	Hipertensão estágio 2 (VALOR mmHg).
	S >= 140 OU D >= 90	Hipertensão estágio 1 (VALOR mmHg).
	S >= 120 OU D >= 80	Pré-hipertensão (VALOR mmHg).
	outros	Pressão arterial normal (VALOR mmHg).

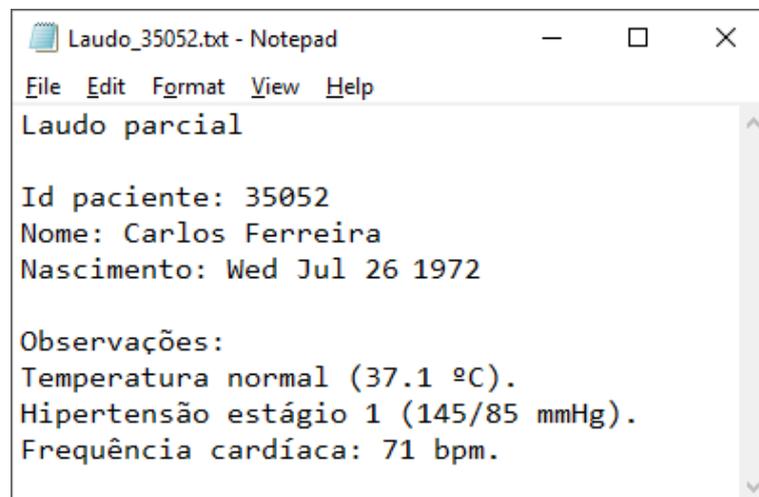
Fonte: Elaborado pelo autor.

De acordo com o valor retornado para cada tipo de característica, é selecionada uma frase correspondente dentro de um conjunto de frases pré-determinadas. Essas sentenças

podem ser definidas por profissionais da saúde, que determinam as regras para a sua exibição. No protótipo desenvolvido as regras foram definidas com base na literatura. O indicador relacionado à temperatura corporal foi formulado a partir da definição de Sund-Levander, Forsberg e Wahren (2002). Já as regras referentes à pressão arterial foram descritas por American Heart Association<sup>31</sup>. A Tabela 10 apresenta as características, regras consideradas e frases correspondentes, sendo que a literal “VALOR” é substituída pela medida da característica. Em relação à pressão sanguínea, “S” representa a medida da pressão sistólica e “D” representa a medida da pressão diastólica.

A saída gerada pelo protótipo é um arquivo de texto contendo duas seções. Na primeira delas é apresentada a identificação do paciente, seu nome e data de nascimento. Na segunda seção estão as frases correspondentes às condições disponíveis no repositório de características, definidas a partir das regras especificadas. A Figura 43 ilustra o conteúdo de um arquivo gerado pela aplicação.

**Figura 43: Exemplo de laudo preliminar gerado pelo segundo protótipo.**



Fonte: Elaborado pelo autor.

#### 5.2.4 Coleta de dados

As informações de entrada do segundo protótipo são imagens de tomografia computadorizada do tórax, cuja obtenção foi descrita no detalhamento do primeiro protótipo, e dados provenientes de sensores de temperatura corporal, pressão sanguínea e frequência cardíaca. Os valores de pressão sanguínea foram retirados da tabela disponibilizada por Cox e Snell (1981), que apresenta 15 medidas da pressão sistólica e diastólica feitas antes e depois do consumo de medicação. Já os dados relacionados à temperatura corporal e frequência cardíaca foram publicados por Shoemaker (1996). Nesse conjunto, composto por 130 aferições, a temperatura é expressa utilizando a escala Fahrenheit (°F). Para uso no protótipo os valores foram convertidos para Celcius (°C).

<sup>31</sup> American Heart Association – [http://www.heart.org/HEARTORG/Conditions/HighBloodPressure/KnowYourNumbers/Understanding-Blood-Pressure-Readings\\_UCM\\_301764\\_Article.jsp](http://www.heart.org/HEARTORG/Conditions/HighBloodPressure/KnowYourNumbers/Understanding-Blood-Pressure-Readings_UCM_301764_Article.jsp)

### 5.2.5 Cenário de uso

Da mesma forma que no primeiro protótipo, a seguir é descrito um cenário de uso para a aplicação desenvolvida e, posteriormente, são apresentados os detalhes do seu funcionamento.

Descrição do cenário:

Ao receber o diagnóstico de que possui pontos de calcificação na região cardíaca, o paciente João da Silva sentiu um mal subido e ficou internado no hospital, em observação. A enfermeira responsável conectou ao paciente dispositivos que fazem a aferição de três sinais vitais: pressão, temperatura e frequência cardíaca. Ao chegar no seu turno, o médico responsável solicitou à enfermeira um relatório geral sobre o estado de saúde do paciente, com o objetivo de avaliar a condição em que ele se encontrava naquele momento. A enfermeira encarregada executou, no sistema do hospital, a geração de um laudo preliminar do estado de saúde do paciente e entregou ao médico solicitante.

Como esse paciente já possui cadastro no hospital não é necessária a etapa de inserção dos seus dados pessoais no sistema. Relembrando, o identificador gerado para o João da Silva é 30054. Após estar acomodado na ala onde ficam os pacientes em observação, já com os sinais vitais sendo acompanhados pela equipe de enfermeiras, foi inserido no sistema um registro contendo a temperatura, pressão e frequência cardíaca do paciente João. A Figura 44 apresenta o *DiagnosticReport* correspondente à operação descrita.

**Figura 44: Representação do *DiagnosticReport* com os sinais vitais do paciente João.**

```
<DiagnosticReport xmlns="http://hl7.org/fhir">
  <id value="45055"/>
  <meta>
    <versionId value="1"/>
    <lastUpdated value="2017-09-17T17:30:20.535-03:00"/>
  </meta>
  <status value="partial"/>
  <subject>
    <reference value="Patient/30054"/>
  </subject>
  <effectiveDateTime value="2017-09-17T17:30:20-03:00"/>
  <issued value="2017-09-17T17:30:20.504-03:00"/>
  <result>
    <reference value="Observation/45052"/>
  </result>
  <result>
    <reference value="Observation/45053"/>
  </result>
  <result>
    <reference value="Observation/45054"/>
  </result>
</DiagnosticReport>
```

Fonte: Elaborado pelo autor.

O *DiagnosticReport* contém, na tag “*subject*”, a referência para o paciente João. Nos resultados estão as três referências para recursos do tipo *Observation*, sendo uma para cada sinal vital, cujos identificadores são 45052, 45053 e 45054. No momento em que foi disparada a inserção das medições correspondentes à cada um dos sinais vitais no repositório, foi executada a seleção, de maneira aleatória, dos valores a serem utilizados, conforme os disponibilizados nas tabelas descritas na seção que detalhou o processo de coleta dos dados para uso no segundo protótipo. O conteúdo, na forma de XML, dos recursos citados está agrupado e exibido na Figura 45.

**Figura 45: Agrupamento dos recursos com os sinais vitais vinculados ao paciente.**

```
<Observation xmlns="http://hl7.org/fhir">
  <id value="45052"/>
  <valueQuantity>
    <value value="36.4"/>
    <unit value="°C"/>
    <system value="http://unitsofmeasure.org"/>
    <code value="Cel"/>
  </valueQuantity>
  <comment value="Temperatura corporal"/>
</Observation>

<Observation xmlns="http://hl7.org/fhir">
  <id value="45053"/>
  <valueString value="166/121"/>
  <comment value="Pressão arterial"/>
</Observation>

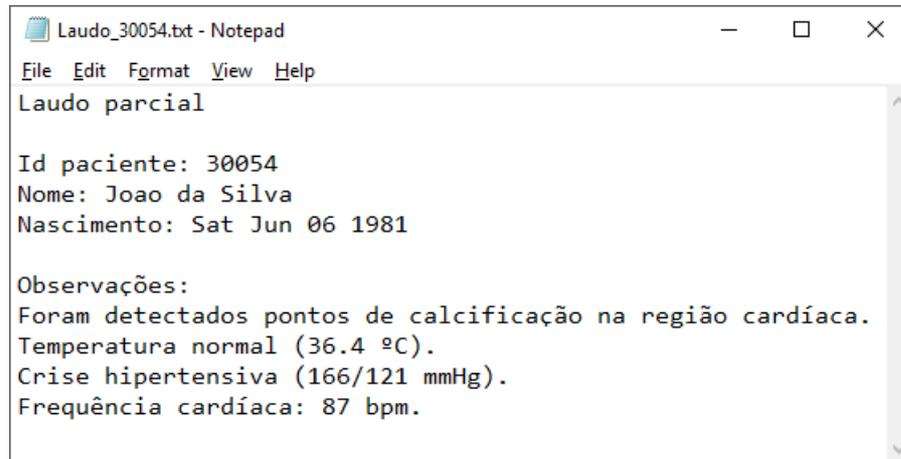
<Observation xmlns="http://hl7.org/fhir">
  <id value="45054"/>
  <valueQuantity>
    <value value="87"/>
  </valueQuantity>
  <comment value="Frequência cardíaca"/>
</Observation>
```

Fonte: Elaborado pelo autor.

Após o médico ter solicitado um relatório com as informações referentes ao estado de saúde do paciente João, a enfermeira utilizou a aplicação de geração de laudo preliminar, inserindo como entrada o identificador do respectivo paciente. A busca por instâncias do recurso *DiagnosticReport* desse paciente retornou dois resultados. O primeiro deles referente ao laudo que indica a existência de pontos de calcificação na região cardíaca, que havia sido inserido no sistema quando o paciente fez seu exame de tomografia computadorizada (primeiro cenário de uso, descrito na seção 5.1.6). A segunda instância é a que foi apresentada anteriormente, que contém as aferições dos sinais vitais.

O conteúdo das duas instâncias foi submetido à aplicação de geração de laudos, que, através das regras definidas pelos especialistas, determina as frases a serem inseridas no documento. O resultado, apresentado na Figura 46, corresponde ao relatório entregue pela enfermeira ao médico solicitante.

**Figura 46: Resultado do laudo preliminar gerado para o paciente João.**



```
Laudo_30054.txt - Notepad
File Edit Format View Help
Laudo parcial

Id paciente: 30054
Nome: Joao da Silva
Nascimento: Sat Jun 06 1981

Observações:
Foram detectados pontos de calcificação na região cardíaca.
Temperatura normal (36.4 °C).
Crise hipertensiva (166/121 mmHg).
Frequência cardíaca: 87 bpm.
```

Fonte: Elaborado pelo autor.

### 5.3 Considerações

Os protótipos construídos, de acordo com a estrutura do modelo proposto, possibilitam o tratamento de três tipos de entrada (imagens, textos e dados provenientes de sensores), provêm o armazenamento dos dados em um repositório padronizado, com o uso da especificação HL7 FHIR e instanciam duas aplicações, uma que realiza o processo de verificação de laudos médicos e outra que executa a geração de um laudo preliminar. O sistema formado pelos dois protótipos é composto por elementos existentes em todas as camadas do modelo apresentado, demonstrando assim o papel de cada componente, sua importância para o sistema e o relacionamento com os demais blocos que completam a estrutura especificada.



## 6 AVALIAÇÃO

Neste capítulo são detalhados os processos de avaliação aplicados ao modelo proposto, bem como os resultados obtidos. O modelo foi submetido a dois tipos de avaliação. O primeiro foi realizado com base na descrição dos protótipos e dos cenários de uso correspondentes, conforme descrito, em parte, no Capítulo 5 e ampliado na seção 6.1. O segundo tipo de avaliação foi realizado através de questionários aplicados a profissionais da saúde e a profissionais da computação, possibilitando avaliar a estrutura, a importância e as contribuições do modelo. A participação de dois perfis de entrevistados permitiu avaliar as contribuições percebidas na área de desenvolvimento de software e na área de uso de aplicações de apoio clínico.

### 6.1 Cenários de uso

Esta seção apresenta a avaliação de funcionalidade baseada em cenários, na qual são definidas possíveis situações a que os sistemas construídos com base no modelo proposto podem ser submetidos. O detalhamento do cenário e a análise das alternativas disponibilizadas a partir do modelo permitem identificar os benefícios de seu emprego, bem como suas limitações e pontos de melhoria.

Foram descritos dois cenários de uso, um para cada um dos protótipos desenvolvidos, nas seções correspondentes (5.1.6 e 5.2.5). Estes cenários apresentam a possibilidade de emprego do modelo em sistemas de apoio clínico, que fazem uso de múltiplos dados de entrada e que utilizam métodos já publicados por outros autores. Nas próximas seções são apresentados outros dois cenários de uso, que reforçam a característica modular e a propriedade expansiva do modelo.

#### 6.1.1 Substituição/atualização do padrão de representação

O cenário descrito abaixo apresenta uma situação em que o padrão de representação precisa ser substituído, em um sistema composto por diversas aplicações e funções diferentes.

Descrição do cenário:

O Hospital Universitário possui todas as suas aplicações computacionais de apoio clínico concentradas em um único sistema, que é utilizado em todos os setores. Para garantir a padronização e a organização das informações, o setor responsável pelo seu desenvolvimento optou por empregar o padrão de representação e armazenamento openEHR. Após um grande período de tempo, o sistema utilizado no hospital já era composto por um elevado número de aplicações, com diferentes finalidades, como apoio ao diagnóstico, funções administrativas, controle de medicamentos e de pacientes, entre outras. Recentemente, a universidade que mantém o hospital firmou um acordo internacional relacionado ao uso do padrão HL7 FHIR, informando ao setor de desenvolvimento que todo o sistema do hospital deveria ser migrado para este padrão. É fundamental que as informações existentes

atualmente não sejam perdidas e que todas as aplicações sejam migradas no menor espaço de tempo possível.

Como todas as aplicações foram construídas utilizando como base o modelo proposto nesta pesquisa, o setor responsável pelo sistema apenas precisou determinar as equivalências entre o padrão em uso atualmente (openEHR) e o padrão que o iria substituir (FHIR). Após essa ação, somente as interfaces de entrada e de saída necessitaram ser modificadas, resultando em uma adaptação muito menos complexa, menos custosa e menos propensa a erros do que a alteração de todas as aplicações que compõem o sistema. Isso só foi possível pois são os elementos de interface que contém as rotinas de acesso direto ao repositório de características a partir das entradas (rotinas de extração de características) e a partir das saídas (aplicações).

Os mesmos métodos de acesso aos dados, que foram modificados nas interfaces de entrada e de saída, além de serem utilizados pelas rotinas de extração e pelas aplicações, poderiam ser utilizados para a migração dos registros já armazenados. Ou seja, a partir da interface de saída antiga seria possível acessar as informações armazenadas na base que utiliza o padrão openEHR e, com o uso da interface de entrada nova, seriam inseridas na base que utiliza a especificação HL7 FHIR. Dessa forma, as informações atualmente existentes seriam transferidas para o novo padrão e o sistema continuaria funcionando com a nova especificação.

Além da substituição completa de um padrão de representação por outro, as interfaces de entrada e saída também podem simplificar a adaptação do sistema para eventuais atualizações que podem ser criadas para o padrão empregado. Por exemplo, caso o recurso *Patient* do HL7 FHIR seja dividido em outros dois recursos, somente as interfaces de entrada e saída precisarão ser adaptadas para que todo o sistema seja preparado para o novo formato.

### 6.1.2 Ampliação de funcionalidades

O cenário descrito abaixo apresenta uma situação em que se deseja adicionar uma nova aplicação a um sistema já desenvolvido e em uso, reaproveitando os dados existentes no repositório de características.

Descrição do cenário:

Uma nova determinação do Ministério da Saúde obriga que todos os hospitais passem a considerar, através do seu sistema de informação, os dados disponíveis sobre cada um dos pacientes ao prescrever tratamentos à base de medicamentos. Essa ação, segundo o próprio ministério, tem como objetivo reduzir o número de problemas de saúde decorrentes do uso de medicações que podem conter substâncias as quais as pessoas, eventualmente, são alérgicas.

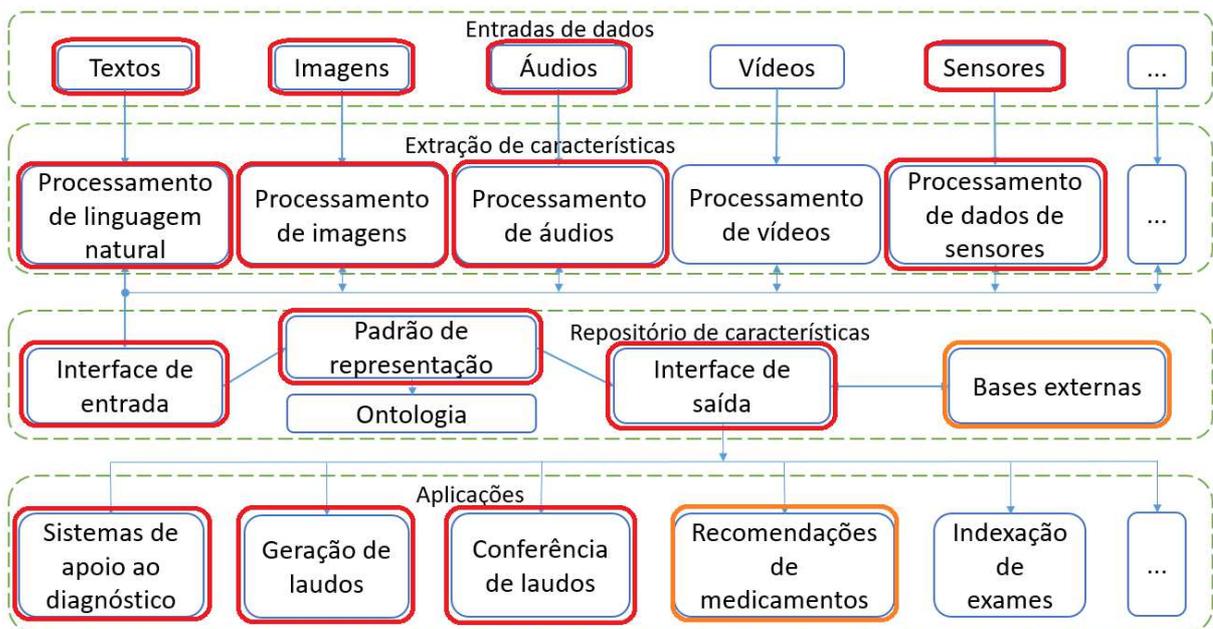
Como o sistema de prescrição de medicamentos terá que ser modificado, a direção do Hospital Universitário determinou que, além de avaliar as substâncias existentes em cada medicamento, sejam considerados também os riscos referentes às interações entre medicações diferentes que são utilizadas de forma concomitante. O sistema atualmente em uso no hospital já possui dados dos pacientes, onde estão indicadas suas alergias, e as receitas de medicamentos entregues a cada paciente. Após avaliar as informações disponíveis no repositório de características, a equipe responsável pelo sistema em uso no hospital identificou que nenhum dos componentes precisa ser modificado. Para atender às definições

da direção era necessário apenas desenvolver dois componentes: Bases externas e Recomendação de medicamentos.

O primeiro elemento citado é necessário pois as informações referentes à composição de cada medicamento estarão disponíveis em um repositório centralizado, mantido pelos fabricantes dos medicamentos, e não na base de dados do hospital. Dessa maneira, é fundamental que o sistema do hospital consiga acessar informações disponibilizadas por terceiros, e o componente denominado Bases externas é o que possibilita essa ação.

Já o segundo elemento representa o sistema de recomendação de medicamentos, cuja finalidade é determinar a medicação mais adequada a cada um dos pacientes, levando em consideração suas condições atuais (determinadas pelos diferentes laudos sobre o seu estado existentes no repositório de características), as informações disponibilizadas pelo paciente sobre suas alergias (definição do Ministério da Saúde) e, por fim, as substâncias que o paciente já está consumindo em seu tratamento corrente (definição da direção do hospital). A Figura 47 apresenta a estrutura atual do sistema utilizado no hospital, em vermelho, e os elementos a serem desenvolvidos para atender a nova demanda, em laranja.

**Figura 47: Composição do sistema atual (vermelho) e elementos adicionados (laranja).**



Fonte: Elaborado pelo autor.

## 6.2 Questionários

A avaliação do modelo foi formulada com base na proposta de Davis (1989), que foi posteriormente estendida por Venkatesh e Davis (2000). Os questionários foram aplicados a profissionais da saúde e a profissionais da computação, após assistirem à apresentação do modelo proposto e dos protótipos desenvolvidos. O objetivo é verificar a aceitabilidade do modelo a partir das percepções desses profissionais em relação a sua facilidade de uso, utilidade e contribuições.

Foram definidos dois formulários, um para cada área (computação e saúde). Cada questionário é composto por uma série de afirmações sobre o tema de pesquisa, o modelo e os elementos que o compõem. O profissional que está respondendo escolhe a opção que reflete, com mais precisão, a sua opinião sobre cada uma das questões, seguindo a escala apresentada por Likert (1932). As seções seguintes detalham os processos de avaliação e os resultados obtidos.

### 6.2.1 Profissionais da saúde

A etapa de avaliação do modelo direcionada aos profissionais da saúde ocorreu em um encontro presencial, que envolveu um total de 12 participantes. Nessa ocasião, a motivação da pesquisa foi apresentada, os trabalhos relacionados foram citados, o modelo foi detalhado (com ênfase para as camadas de entradas e aplicações) e os protótipos foram demonstrados. Do grupo de participantes, quatro atuam como enfermeiros, quatro são estudantes pesquisadores, três são professores de ensino superior e um é biomédico. A faixa etária variou entre 20 e 44 anos. O requisito para a participação nessa etapa da avaliação é conhecer aplicações de apoio clínico. Após a apresentação, os convidados responderam um questionário contendo seis questões objetivas (obrigatórias) e três questões dissertativas (opcionais).

As três primeiras questões objetivas visam verificar o atual cenário referente ao uso de aplicações de apoio clínico que fazem uso, de maneira integrada, de múltiplas fontes de dados, buscando confirmar as motivações que levaram ao desenvolvimento desta pesquisa. A Tabela 11 apresenta essas questões e os resultados obtidos na avaliação.

**Tabela 11: Cenário atual de uso de aplicações de apoio que empregam dados heterogêneos.**

<b>Questão</b>	<b>Discordo totalmente</b>	<b>Discordo parcialmente</b>	<b>Não concordo nem discordo</b>	<b>Concordo parcialmente</b>	<b>Concordo totalmente</b>
1. Sinto que há carência de aplicações que fazem uso integrado de informações provenientes de diferentes fontes.	0 (0%)	0 (0%)	3 (25%)	3 (25%)	6 (50%)
2. O uso de dados de diferentes fontes pode melhorar os resultados fornecidos pelas aplicações de apoio clínico existentes.	0 (0%)	0 (0%)	1 (8,3%)	2 (16,7%)	9 (75%)
3. O uso de dados de diferentes fontes pode viabilizar o desenvolvimento de novas aplicações de apoio clínico.	0 (0%)	0 (0%)	0 (0%)	2 (16,7%)	10 (83,3%)

Fonte: Elaborado pelo autor.

Após a questão objetiva número 2, foi adicionada uma questão dissertativa, solicitando aos avaliadores para que fossem citados tipos de exames que considerem relevantes para

aplicações de apoio clínico. Os tipos de dados mais citados foram exames laboratoriais, exames por imagem (ressonância magnética, tomografia computadorizada, radiografia), exame físico, sinais vitais e prescrições. Todos os tipos de informações indicados podem ser utilizados em sistemas construídos com base no modelo proposto.

Outra questão dissertativa foi inserida após a questão objetiva número 3, disponibilizando um espaço para o participante inserir sugestões de sistemas computacionais que empregam dados heterogêneos. Dentre os itens citados, destaca-se uma sugestão de aplicação que considera as condições disponíveis do paciente e sugere a conduta terapêutica com maior probabilidade de sucesso. Também foi indicada a necessidade de considerar, além de exames laboratoriais e por imagem, os hábitos de vida dos pacientes. Como exemplificado no modelo, um dos avaliadores apresentou também a possibilidade de construção de uma aplicação capaz de sugerir medicamentos de acordo com as características do paciente.

As outras três questões objetivas abordam a utilidade e as contribuições do modelo, sendo uma relacionada ao elemento que corresponde à ontologia, na camada do repositório de características. A Tabela 12 apresenta essas questões e as opiniões dos participantes.

**Tabela 12: Opinião dos profissionais da saúde sobre a utilidade e divulgação do modelo.**

Questão	Discordo totalmente	Discordo parcialmente	Não concordo nem discordo	Concordo parcialmente	Concordo totalmente
6. Na minha opinião, o modelo proposto é útil.	0 (0%)	0 (0%)	0 (0%)	5 (41,7%)	7 (58,3%)
7. O uso da ontologia no modelo proposto simplifica o entendimento e a visualização do conteúdo do repositório de características.	0 (0%)	0 (0%)	0 (0%)	4 (33,3%)	8 (66,7%)
8. Se eu tiver oportunidade, irei divulgar o modelo proposto.	0 (0%)	0 (0%)	1 (8,3%)	3 (25%)	8 (66,7%)

Fonte: Elaborado pelo autor.

O último item existente no questionário é um espaço para que os participantes pudessem inserir sugestões para o modelo apresentado. Houve apenas uma colocação, relacionada ao protótipo de geração de laudo preliminar. O profissional indicou que o protótipo poderia ser complementado, de modo a evitar que o laudo fosse utilizado sem que cada observação fosse, individualmente, revisada. O objetivo seria reduzir a possibilidade de oficialização de informações que não foram avaliadas pelo responsável.

## 6.2.2 Profissionais da computação

A etapa de avaliação do modelo direcionada aos profissionais da computação ocorreu em três encontros presenciais, envolvendo um total de 35 participantes. Nesses encontros, a motivação da pesquisa foi apresentada, o modelo foi detalhado (com ênfase para a camada do repositório de características) e os protótipos foram demonstrados. Do grupo de participantes,

vinte e dois atuam como desenvolvedores de software, cinco são analistas de sistemas, quatro são professores de ensino superior e quatro são estudantes pesquisadores. A faixa etária variou entre 18 e 53 anos. O requisito para a participação nessa etapa da avaliação é ter grande familiaridade com o desenvolvimento de aplicações computacionais. Após a apresentação, os convidados responderam um questionário contendo oito questões objetivas (obrigatórias) e uma questão dissertativa (opcional).

As cinco primeiras questões objetivas visam verificar se o modelo descrito é composto pelos elementos necessários, se ele é capaz de preencher as lacunas de pesquisa identificadas e se consegue responder à questão de pesquisa apresentada. A Tabela 13 contém essas questões e os resultados obtidos na avaliação.

**Tabela 13: Avaliação da estrutura e das contribuições do modelo proposto.**

Questão	Discordo totalmente	Discordo parcialmente	Não concordo nem discordo	Concordo parcialmente	Concordo totalmente
1. O modelo proposto pode simplificar a construção e a ampliação de funcionalidades em aplicações de apoio clínico.	0 (0%)	0 (0%)	1 (2,9%)	9 (25,7%)	25 (71,4%)
2. O modelo proposto apresenta uma estrutura mais completa do que as utilizadas atualmente na construção de aplicações de apoio clínico.	0 (0%)	0 (0%)	7 (20%)	18 (51,4%)	10 (28,6%)
3. A composição do modelo proposto (elementos e relacionamentos) é clara.	0 (0%)	0 (0%)	5 (14,3%)	12 (34,3%)	18 (51,4%)
4. A estrutura do modelo proposto viabiliza a adição e/ou substituição de elementos de maneira independente (sem que seja necessário alterar os demais elementos já utilizados).	0 (0%)	0 (0%)	2 (5,7%)	14 (40%)	19 (54,3%)
5. A existência das interfaces de entrada e saída no modelo proposto permite o uso de pesquisas já desenvolvidas com mais facilidade.	0 (0%)	0 (0%)	1 (2,9%)	9 (25,7%)	25 (71,4%)

Fonte: Elaborado pelo autor.

Após o primeiro grupo de questões objetivas, que fazem referência às especificações e aos benefícios que podem ser obtidos através do uso do modelo proposto, os avaliadores responderam ao segundo grupo de questões objetivas, que abordam a utilidade e as

contribuições do modelo, sendo uma relacionada ao elemento que corresponde à ontologia, na camada do repositório de características. A Tabela 14 apresenta essas questões e as opiniões dos participantes.

**Tabela 14: Opinião dos profissionais da computação sobre a utilidade do modelo.**

Questão	Discordo totalmente	Discordo parcialmente	Não concordo nem discordo	Concordo parcialmente	Concordo totalmente
6. Na minha opinião, o modelo proposto é útil.	0 (0%)	0 (0%)	1 (2,9%)	4 (11,4%)	30 (85,7%)
7. O uso da ontologia no modelo proposto simplifica o entendimento e a visualização do conteúdo do repositório de características.	0 (0%)	0 (0%)	3 (8,6%)	8 (22,8%)	24 (68,6%)
8. Se eu for desenvolver um sistema de apoio clínico, irei utilizar o modelo proposto.	0 (0%)	1 (2,9%)	9 (25,7%)	14 (40%)	11 (31,4%)

Fonte: Elaborado pelo autor.

O último item existente no questionário é um espaço para que os participantes pudessem inserir sugestões para o modelo apresentado. O destaque nessa questão foram as colocações relacionadas ao uso da ontologia como forma alternativa para visualização da estrutura e do conteúdo do repositório de características. Um dos avaliadores indicou que é possível que existam outros formatos de representação que podem ser mais compreensíveis para outros públicos. Dessa forma seria interessante explorar também outras maneiras, mas não foram citados exemplos.

### 6.3 Considerações

Os métodos de avaliação empregados visam analisar diferentes aspectos do modelo, como possibilidades de uso, aderência e complexidade de adaptação de métodos já existentes, benefícios obtidos em relação à sua modularidade e expansibilidade, relevância, aceitação, entre outras características. Tanto a avaliação através de cenários de uso como a avaliação feita com a participação de profissionais viabilizam a identificação de aspectos positivos e também possíveis melhorias que podem ser feitas no modelo.

Os quatro cenários de uso descritos nas seções anteriores detalham o comportamento de um sistema construído com base no modelo proposto em relação à diferentes situações. O primeiro cenário corresponde à aplicação de conferência de laudos (seção 5.1.6), demonstrando o uso integrado de dados provenientes de diferentes fontes (textos e imagens). Nesse cenário é descrito o funcionamento prático dos elementos desenvolvidos, com destaque para os componentes existentes na camada do repositório de características. O segundo cenário (seção 5.2.5) descreve o uso da aplicação de geração de laudo preliminar, em que foi feita a adição de um novo tipo de dado de entrada ao sistema desenvolvido no primeiro protótipo: informações provenientes de sensores. Nesse caso, é demonstrada a característica

modular, de modo que foi inserido um novo tipo de entrada e apenas o componente de interface correspondente precisou ser modificado, além da característica expansiva do modelo, já que o número de funcionalidades do sistema foi ampliado de forma simples, reaproveitando informações já existentes no repositório.

O terceiro cenário de uso (seção 6.1.1) tem como foco a modularidade do modelo, uma vez que todo o padrão de armazenamento e representação utilizado necessita ser modificado. Esse processo é simplificado através do emprego das interfaces de entrada e de saída. Se não houvessem esses elementos, o volume de modificações que seriam necessários no sistema para adaptá-lo à nova especificação seria muito maior. Por fim, o último cenário de uso (seção 6.1.2) enfatiza a simplicidade de se expandir funcionalidades em sistemas desenvolvidos com base no modelo proposto. Nesse exemplo, não foi necessário processar uma nova entrada. Foram utilizadas as informações já existentes no repositório de características e dados complementares são acessados através do elemento de ligação com bases externas.

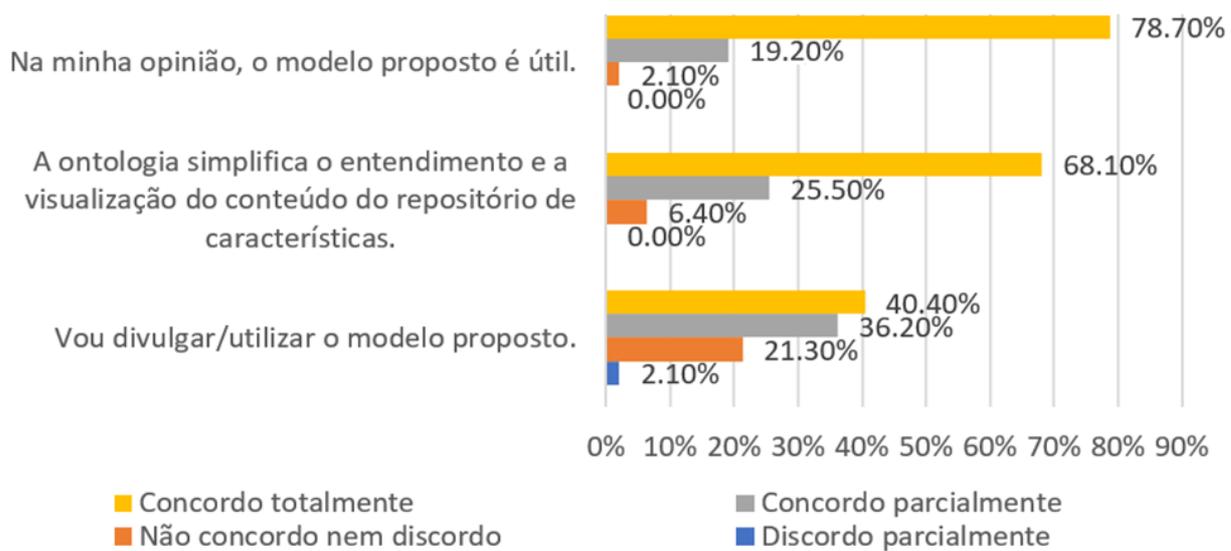
Dentre todos os elementos existentes no modelo proposto, apenas o correspondente à ontologia não foi empregado nos protótipos e cenários descritos. As percepções sobre os benefícios de uso desse componente foram confirmadas através dos questionários, como será apresentado adiante.

A maior parte dos profissionais da área da saúde que respondeu o questionário concorda, de maneira parcial ou total, que há carência de aplicações que utilizam múltiplas fontes de dados (75%), que o uso de dados heterogêneos pode trazer melhorias para os sistemas existentes (91,7%) e que novas aplicações que integram diferentes informações podem ser desenvolvidas (100%). Esses percentuais demonstram que há um grande espaço para a evolução dos sistemas de apoio clínico que utilizam dados heterogêneos.

Um grande percentual dos profissionais da computação que participou da avaliação concorda, de maneira parcial ou total, que o modelo pode simplificar o desenvolvimento e a ampliação de sistemas de apoio clínico (97,1%) e que ele apresenta uma estrutura mais completa do que as utilizadas atualmente (80%). Também foi indicado que o modelo está definido de forma clara (85,7%) e que a sua estrutura é modular e expansiva, viabilizando a substituição e/ou adição de elementos de maneira independente (94,3%). Por fim, quase todos participantes dessa área entendem que a existência das interfaces de entrada e de saída viabilizam a adaptação de pesquisas já desenvolvidas de maneira mais simples (97,1%).

As questões do último grupo de ambos os questionários podem ser consideradas equivalentes, possibilitando a junção dos resultados. A Figura 48 ilustra esse agrupamento. Através dessa Figura pode-se identificar que o modelo proposto é considerado útil (97,9%) e que os participantes podem vir a utilizá-lo ou divulgá-lo (76,6%). Na opinião de 93,6% dos entrevistados, a existência de uma maneira alternativa para visualizar os dados, através de ontologias, simplifica o acesso e o entendimento do conteúdo armazenado no repositório de características, em comparação com a visualização em XML ou JSON do padrão de representação utilizado nos protótipos (HL7 FHIR).

**Figura 48: Resultado do agrupamento das respostas equivalentes nos grupos.**



Fonte: Elaborado pelo autor.



## 7 CONCLUSÃO

Este trabalho apresentou um modelo que pode ser utilizado como base para o desenvolvimento de aplicações de apoio clínico baseado na obtenção e disponibilização, de maneira estruturada, de características a partir do processamento de dados heterogêneos. Pretende-se, com o uso de diferentes fontes de informações, viabilizar a criação de sistemas que farão uso dos dados extraídos de maneira complementar, melhorando aspectos como eficiência, precisão e personalização das atividades clínicas.

A estrutura proposta é composta por uma série de elementos definidos, distribuídos em quatro camadas: entradas de dados, extração de características, repositório de características e aplicações. A partir da análise do estado-da-arte de sistemas computacionais da área médica, que fazem uso de múltiplas fontes de dados de entrada, foi possível identificar as lacunas de pesquisa que foram consideradas e, conseqüentemente, definir os componentes necessários para preenchê-las.

Foram desenvolvidos dois protótipos de aplicações de apoio clínico que fazem uso, de maneira integrada, de informações extraídas a partir do processamento de diferentes tipos de dados. O primeiro executa a conferência de laudos médicos utilizando como entrada imagens de tomografia computadorizada e os textos dos laudos correspondentes. Já o segundo protótipo é responsável pela geração do laudo preliminar do estado de saúde de um paciente, utilizando como entrada imagens médicas e dados provenientes de sensores. Em ambos os casos cada elemento foi implementado de maneira independente, de forma integral pelo autor ou com base em aplicações já validadas e apresentadas na literatura. Isso demonstra a característica modular e expansível da estrutura proposta, preenchendo assim a terceira lacuna de pesquisa apontada na avaliação dos trabalhos relacionados. Dessa forma, elementos podem ser substituídos e novas funcionalidades podem ser adicionadas ao sistema sem que sejam necessárias alterações nos demais componentes. O uso de diferentes tipos de dados como entrada e o emprego de um padrão de representação reconhecido, como o utilizado nos protótipos (HL7 FHIR), preenchem, respectivamente, as demais lacunas de pesquisas descritas. Também é indicada a possibilidade de comunicação com bases externas, tanto para viabilizar o uso de informações que estão disponíveis em outros repositórios como para disponibilizar os dados armazenados no repositório local para serem empregados em aplicações de terceiros.

A avaliação do modelo foi feita a partir de dois métodos: cenários de uso e questionários respondidos por profissionais. O desenvolvimento dos protótipos permitiu verificar a capacidade de utilização e integração dos itens que formam o modelo. Isso também viabilizou a realização de provas de conceito através dos quatro cenários de uso formulados. Por meio das respostas fornecidas pelos protótipos em relação ao detalhamento de cada um dos cenários, foi possível coletar indicações no sentido de que o modelo atendeu a todas as expectativas colocadas pelos seus objetivos, além de possibilitar a percepção das suas contribuições e eventuais fragilidades. A avaliação dos profissionais envolveu 47 participantes, sendo 12 atuantes na área da saúde e 35 na área da computação. Através das opiniões dos entrevistados da área da saúde foi possível confirmar que há carência de aplicações computacionais que fazem uso, de maneira integrada, de múltiplas fontes de entrada, e que o emprego de diferentes tipos de dados pode contribuir para a obtenção de avanços nas aplicações existentes e, conseqüentemente, melhorar a qualidade do atendimento prestado aos pacientes. Já os profissionais da computação indicaram sua concordância quanto aos benefícios de uso do modelo na construção de sistemas desse tipo, principalmente em

relação à sua modularidade e à sua capacidade de expansão de funcionalidades. Para ambos os grupos de profissionais o uso de ontologia foi considerado positivo, sendo uma maneira alternativa para visualização dos dados e da organização do repositório de características, simplificando o acesso e o entendimento das informações disponíveis.

A principal contribuição deste trabalho é o detalhamento de uma estrutura modular e expansível, que possibilita o emprego de características extraídas a partir do processamento de diferentes tipos de entrada, de maneira integrada, em aplicações de apoio clínico. O modelo disponibiliza uma estrutura de referência para auxiliar na construção e integração de aplicações médicas, simplificando o emprego de diferentes tipos de dados como entrada, com ênfase na ampliação de funcionalidades e no uso de pesquisas já desenvolvidas. A existência de uma estrutura como essa pode fomentar a utilização de pesquisas com diferentes finalidades de maneira complementar, tornando-as mais completas e de maior abrangência. Essa possibilidade aumenta a integração entre sistemas já existentes e o desenvolvimento de novas aplicações, contribuindo assim para a ampliação do apoio da computação na área clínica e viabilizando a evolução do estado-da-arte.

Os benefícios obtidos através do emprego do modelo são acentuados no momento em que é necessário integrar diferentes aplicações ou ampliar o número de recursos de um sistema existente, principalmente se houverem métodos já desenvolvidos que precisam ser adicionados. A estrutura apresentada define como relacionar os componentes existentes com os que podem ser incorporados ao sistema. Nos trabalhos relacionados essas possibilidades não são observadas. Cada pesquisa trata de uma maneira diferente as informações obtidas e os resultados, não havendo discussões sobre como podem ser flexibilizados e compartilhados.

Em comparação ao estado-da-arte, o maior diferencial do modelo proposto é a definição da camada denominada Repositório de características, composta pelas interfaces de entrada e de saída, padrão de representação, comunicação com bases externas e ontologia. As interfaces são os elementos de maior destaque, uma vez que são os responsáveis pela compatibilização e interlocução entre os componentes de extração das características, repositório dos dados e aplicações, removendo a complexidade relacionada à especificação do padrão escolhido para armazenar as informações. O elemento que representa a ontologia é outro diferencial, indicando a possibilidade de acessar e visualizar o conteúdo do repositório de uma maneira alternativa, sem que o usuário tenha necessidade de conhecer detalhes do padrão de representação.

Em relação à questão de pesquisa e às lacunas identificadas nos trabalhos relacionados, conclui-se que as propriedades relevantes de um modelo baseado em processamento de dados heterogêneos para aplicações de apoio clínico abrangem os diferentes elementos que compõem a estrutura apresentada, em conjunto com algumas características relacionadas ao seu uso, tais como:

- a) Entradas e extração de características: São admitidos diferentes tipos de dados (textos, imagens, vídeos, sons, sensores e outros) como entrada, possibilitando a execução de métodos para obter características relevantes em cada um deles;
- b) Integrações: A disponibilização de características que podem ser relacionadas através de um padrão de representação reconhecido e das interfaces de entrada e saída permite o uso, de maneira integrada, de informações extraídas a partir de diferentes tipos de dados e a interoperabilidade entre aplicações já desenvolvidas. O modelo indica também a possibilidade de trocar informações com bases externas, que podem ser acessadas em outros sistemas ou locais;

- c) Expansão e adaptação: A estrutura modular e expansível do modelo permite a alteração, adição ou substituição dos elementos que o formam por outros equivalentes, sem que seja necessário modificar os demais componentes já existentes no sistema.

No modelo não são detalhadas as especificações técnicas referentes à implementação de cada um dos componentes, principalmente aqueles relacionados aos processos de extração das características. Isso acontece porque essas definições dependem fortemente do tipo de dado a ser utilizado, do padrão em que ele se encontra e das características que se deseja obter. As aplicações também não são definidas, de modo que foram citados apenas alguns exemplos para demonstrar o uso do modelo.

O modelo apresentado possui algumas oportunidades para trabalhos futuros, como:

- Criação de um arquivo de parametrização para as interfaces de entrada e saída, possibilitando a definição de configurações que determinam o comportamento desses elementos. Um exemplo de configuração pode ser o padrão de representação a ser utilizado. Dessa forma, seria possível empregar a mesma implementação das interfaces em sistemas diferentes que usam padrões de representação diferentes;
- Pode ser necessário utilizar dois, ou mais, padrões de representação em paralelo, caso não seja possível, por exemplo, representar uma informação específica em um determinado padrão. Nesse caso, seria necessário preparar as interfaces para trabalharem com mais de um padrão de representação ao mesmo tempo. Isso pode ser alcançado com o uso de métodos diferentes de acordo com o padrão que se deseja utilizar ou com a modificação dos métodos existentes, passando a receber, por parâmetro, o padrão a ser utilizado;
- O desenvolvimento de uma maneira para criação dinâmica dos métodos de manipulação dos dados disponíveis nas interfaces de entrada e de saída é outro possível trabalho futuro. Foram implementados, no protótipo, apenas os métodos necessários para a avaliação do modelo. A existência desse recurso poderia simplificar a adição de elementos que manipulam tipos de dados ainda não existentes no repositório, tornando o modelo mais genérico;
- Pode ser desenvolvida uma maneira de gerar, dinamicamente, a ontologia que apresenta a estrutura, o relacionamento e os dados armazenados no repositório de características;
- Outro trabalho futuro aborda a definição das especificações do componente Bases externas, padronizando esse elemento de maneira semelhante às interfaces de entrada e de saída.



## REFERÊNCIAS

- ABACHA, Asma B.; ZWEIGENBAUM, Pierre. Medical entity recognition: A comparison of semantic and statistical methods. In: BioNLP 2011, Portland, Oregon. **Proceedings of the 2011 Workshop on Biomedical Natural Language Processing**. Stroudsburg, PA: ACL, 2011, p. 56-64.
- AGAH, Arvin. **Medical Applications of Artificial Intelligence**. Boca Raton, FL: CRC Press, 2013. 460 p.
- ALLWELL-BROWN, Eneimi. **A Comparative Analysis of HL7 FHIR and openEHR for Electronic Aggregation, Exchange and Reuse of Patient Data in Acute Care**. 2016. Dissertação (Mestrado em Informática em Saúde) - Karolinska Institutet, Stockholm University, Stockholm, 2016.
- ALMEIDA, Mauricio B.; BAX, Marcello P. Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. **Ciência da Informação**, Brasília, v. 32, n. 3, p. 7-20, 2003.
- ARAUJO, Tiago V.; PIRES, Silvio R.; BANDIERA-PAIVA, Paulo. Adoção de padrões para Registro Eletrônico em Saúde no Brasil. **Revista Eletrônica de Comunicação, Informação & Inovação em Saúde**, v. 8, n. 4, p. 554-566, 2014.
- ARMANO, Giuliano; FANNI, Francesca; GIULIANI, Alessandro. Stopwords Identification by Means of Characteristic and Discriminant Analysis. In: ICAART 2015, Lisboa, Portugal. **Proceedings of the International Conference on Agents and Artificial Intelligence**. Portugal: SCITEPRESS, 2015, v. 2, p. 353-360.
- BACELAR, Gustavo; CORREIA, Ricardo. **As bases do openEHR**. Porto: VirtualCare, 2015. 42 p.
- BANKMAN, Isaac N. **Handbook of Medical Imaging: Processing and Analysis**. San Diego, CA: Academic Press, 2000. 893 p.
- BEZ, Marta R. **O uso de tecnologia para apoiar a implantação de métodos ativos nos currículos de medicina**. 2011. Tese (Doutorado em Informática na Educação) - Centro Interdisciplinar de Novas Tecnologias na Educação, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, 2011.
- BICK, Eckhard. **The Parsing System "PALAVRAS": Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework**. Aarhus, DK: Aarhus University Press, 2000. 412 p.
- BOSCA, Diego et al. Combining Archetypes with Fast Health Interoperability Resources in Future-proof Health Information Systems. **Digital Healthcare Empowering Europeans: Proceedings of MIE2015**, p. 180-183, 2015.
- BOZKURT, Selen et al. Using automatically extracted information from mammography reports for decision-support. **Journal of Biomedical Informatics**, v. 62, p. 224-231, 2016.

BRASIL. Ministério da Saúde. Portaria nº 2.073, de 31 de agosto de 2011. Regulamenta o uso de padrões de interoperabilidade e informação em saúde para sistemas de informação em saúde no âmbito do Sistema Único de Saúde, nos níveis Municipal, Distrital, Estadual e Federal, e para os sistemas privados e do setor de saúde suplementar. **Diário Oficial da República Federativa do Brasil**, Brasília, 1 set. 2011. Seção 1, p. 63.

BUNTIN, Melinda B. et al. The benefits of health information technology: a review of the recent literature shows predominantly positive results. **Health Affairs**, v. 30, n. 3, p. 464-471, 2011.

CAI, Tianrun et al. Natural Language Processing Technologies in Radiology Research and Clinical Applications. **RadioGraphics**, v. 36, n. 1, p. 176-191, 2016.

CANNY, John. A Computational Approach to Edge Detection. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 8, n. 6, p. 679-698, 1986.

CHENG, Jie-Zhi et al. Computer-Aided diagnosis with deep learning architecture: applications to breast lesions in us images and pulmonary nodules in CT scans. **Scientific Reports**, v. 6, 2016.

CHO, Paul S.; TAIRA, Ricky K; KANGARLOO, Hooshang. Automatic section segmentation of medical reports. In: AMIA 2003. Washington, DC. **AMIA 2003 Symposium Proceedings**. p. 155-159.

CONCI, Aura; AZEVEDO, Eduardo; LETA, Fabiana R. **Computação Gráfica: Teoria e Prática**. Rio de Janeiro, RJ: Elsevier, 2008. 407 p.

COX, David R.; SNELL, Eleanor J. **Applied Statistics: Principles and Examples**. Boca Raton, FL: CRC Press, 1981. 192 p.

DAVIS, Fred D. Perceived usefulness, perceived ease of use, and user acceptance of information technology. **MIS Quarterly**, v. 13, n. 3, p. 319-340, 1989.

DEMNER-FUSHMAN, Dina; CHAPMAN, Wendy W.; MCDONALD, Clement J. What can natural language processing do for clinical decision support?. **Journal of Biomedical Informatics**, v. 42, n. 5, p. 760-772, 2009.

DOUGHERTY, Geoff. **Digital image processing for medical applications**. New York, NY: Cambridge University Press, 2009. 447 p.

EL-DAHSHAN, El-Sayed A. et al. Computer-aided diagnosis of human brain tumor through MRI: A survey and a new algorithm. **Expert Systems with Applications**, v. 41, n. 11, p. 5526-5545, 2014.

FICHMAN, Robert G.; KOHLI, Rajiv; KRISHNAN, Ranjani. Editorial overview—The Role of Information Systems in Healthcare: Current Research and Future Trends. **Information Systems Research**, v. 22, n. 3, p. 419-428, 2011.

FIRMINO, Macedo et al. Computer-aided detection (CADe) and diagnosis (CADx) system for lung cancer with likelihood of malignancy. **BioMedical Engineering Online**, v. 15, n. 2, 2016.

FLORES, Jéferson C. **Detecção Automática de Calcificação Arterial Coronariana em Tomografia Computadorizada**. 2015. 96 f. Trabalho de Conclusão de Curso (Monografia) – Curso de Ciência da Computação, Universidade Feevale, Novo Hamburgo, RS, 2015.

FRIEDMAN, Carol; JOHNSON, Stephen B. Natural language and text processing in biomedicine. In: SHORTLIFFE, Edward H.; CIMINO, James J. (Org.). **Biomedical Informatics**. New York: Springer, 2006. p. 312-343.

FRIEDMAN, Carol; HRIPCSAK, George. Natural language processing and its future in medicine. **Academic Medicine**, v. 74, n. 8, p. 890-895, 1999.

GAO, Wenshuo et al. An improved Sobel edge detection. In: **3rd IEEE International Conference on Computer Science and Information Technology (ICCSIT)**, 2010, Chengdu, CN: IEEE, 2010. p. 67-71.

GARG, Amit X. et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. **Jama**, v. 293, n. 10, p. 1223-1238, 2005.

van GINNEKEN, Bram; SCHAEFER-PROKOP, Cornelia M.; PROKOP, Mathias. Computer-aided diagnosis: how to move from the laboratory to the clinic. **Radiology**, v. 261, n. 3, p. 719-732, 2011.

GINSBURG, Shoshana B. et al. Automated texture-based quantification of centrilobular nodularity and centrilobular emphysema in chest CT images. **Academic Radiology**, v. 19, n. 10, p. 1241-1251, 2012.

GONZALEZ, Rafael C.; WOODS, Richard E. **Processamento Digital de Imagens**. 3. ed. São Paulo, SP: Pearson, 2010. 624 p.

GÜNGÖR, Tunga. Part-of-speech tagging. In: INDURKHYA, Nitin; DAMERAU, Fred J. (Org.). **Handbook of Natural Language Processing**. Boca Raton, FL: Chapman and Hall/CRC, 2010. p. 205-235.

HANDELS, Heinz et al. Viewpoints on medical image processing: from science to application. **Current Medical Imaging Reviews**, v. 9, n. 2, p. 79-88, 2013.

HAZLEHURST, Brian L. et al. CER Hub: An informatics platform for conducting comparative effectiveness research using multi-institutional, heterogeneous, electronic clinical data. **International Journal of Medical Informatics**, v. 84, n. 10, p. 763-773, 2015.

HE, Lifeng; CHAO, Yuyan; SUZUKI, Kenji. A run-based two-scan labeling algorithm. **IEEE Transactions on Image Processing**, v. 17, n. 5, p. 749-756, 2008.

HL7. **Introduction to HL7 Standards**. 2017a. Disponível em: <<http://www.hl7.org/implement/standards/index.cfm>>. Acesso em: 04 abr. 2017.

HL7. **HL7 Version 3 Product Suite**. 2017b. Disponível em: <[http://www.hl7.org/implement/standards/product\\_brief.cfm?product\\_id=186](http://www.hl7.org/implement/standards/product_brief.cfm?product_id=186)>. Acesso em: 06 abr. 2017.

HL7. **CDA® Release 2**. 2017c. Disponível em:

<[http://www.hl7.org/implement/standards/product\\_brief.cfm?product\\_id=7](http://www.hl7.org/implement/standards/product_brief.cfm?product_id=7)>. Acesso em: 07 abr. 2017.

HL7. **Welcome to FHIR®**. 2017d. Disponível em: <<http://hl7.org/fhir/>>. Acesso em: 08 abr. 2017.

HONORATO, Daniel de F. **Metodologia para mapeamento de informações não estruturadas descritas em laudos médicos para uma representação atributo-valor**. 2008. 145 f. Dissertação (Mestrado em Ciências de Computação e Matemática Computacional) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2008.

INGASON, Anton K. et al. A mixed method lemmatization algorithm using a hierarchy of linguistic identities (HOLI). In: GOEBEL, Randy; SIEKMANN, Jörg; WAHLSTER, Wolfgang (Org.). **Advances in Natural Language Processing**. Berlin, Germany: Springer Berlin Heidelberg, 2008. p. 205-216.

JAISWAL, Abeg K.; BANKA, Haider. Epileptic seizure detection in EEG signal with GModPCA and support vector machine. **Bio-Medical Materials and Engineering**, v. 28, n. 2, p. 141-157, 2017.

KANSAGRA, Akash P. et al. Big data and the future of radiology informatics. **Academic Radiology**, v. 23, n. 1, p. 30-42, 2015.

KOCHENBORGER, Lucas R. **Desenvolvimento de um Protótipo de Software para Detecção da Camada Pericárdio**. 2013. 75 f. Trabalho de Conclusão de Curso (Monografia) – Curso de Sistemas de Informação, Universidade Feevale, Novo Hamburgo, RS, 2013.

LIKERT, Rensis. A technique for the measurement of attitudes. **Archives of Psychology**, v. 22, n. 140, p. 5-55, 1932.

LIU, Yueyi I.; RUBIN, Daniel L. The role of informatics in health care reform. **Academic Radiology**, v. 19, n. 9, p. 1094-1099, 2012.

LOBACH, David et al. **Enabling Health Care Decisionmaking Through Clinical Decision Support and Knowledge Management**. Rockville, MD: Agency for Healthcare Research & Quality, 2012. 109 p.

LOPES, Alneu de A.; JORGE, Alípio. Combining rule-based and case-based learning for iterative part-of-speech tagging. In: 5th European Workshop (EWCBR 2000), Trento, Italy. **European Workshop on Advances in Case-Based Reasoning**. Germany: Springer, 2000. p. 26-36.

MAGLOGIANNIS, Ilias et al. Redesigning EHRs and clinical decision support systems for the precision medicine era. In: EANN 2015, Rhodes, Greece. **Proceedings of the 16th International Conference on Engineering Applications of Neural Networks (INNS)**. New York, NY: ACM, 2015.

MARCONI, Marina de A.; LAKATOS, Eva M. **Fundamentos de metodologia científica**. 5. ed. São Paulo, SP: Atlas, 2003. 311 p.

MARTINEZ-COSTA, Catalina; SCHULZ, Stefan. HL7 FHIR: Ontological Reinterpretation of Medication Resources. In: RENDELL, Rebecca et al. (Org.). **Informatics for Health: Connected Citizen-Led Wellness and Population Health**. Amsterdam, Netherlands: IOS Press, 2017. p. 451-455.

MORRISON, James J. et al. Constructing a Computer-Aided Differential Diagnosis Engine from Open-Source APIs. **Journal of Digital Imaging**, v. 29, n.6, p. 654-657, 2016.

MOURÃO, André; MARTINS, Flávio; MAGALHÃES, João. Multimodal medical information retrieval with unsupervised rank fusion. **Computerized Medical Imaging and Graphics**, v. 39, p. 35-45, 2014.

NABIH-ALI, Mohammed; EL-DAHSHAN, El-Sayed A.; YAHIA, Ashraf S. A review of intelligent systems for heart sound signal analysis. **Journal of Medical Engineering & Technology**, p. 1-11, 2017.

NEMA. **The DICOM Standard 2017a**. 2017. Disponível em: <<http://medical.nema.org/standard.html>>. Acesso em: 29 mar. 2017.

OPENEHR. **openEHR Architecture Overview**. 2016. Disponível em: <[http://www.openehr.org/releases/BASE/latest/docs/architecture\\_overview/architecture\\_overview.html](http://www.openehr.org/releases/BASE/latest/docs/architecture_overview/architecture_overview.html)>. Acesso em: 03 abr. 2017.

OPENEHR. **What is openEHR?**. 2017a. Disponível em: <[http://www.openehr.org/what\\_is\\_openehr](http://www.openehr.org/what_is_openehr)>. Acesso em: 02 abr. 2017.

OPENEHR. **Archetype Query Language (AQL)**. 2017b. Disponível em: <<http://www.openehr.org/releases/QUERY/latest/docs/AQL/AQL.html>>. Acesso em: 03 abr. 2017.

OU, Ying; PATRICK, Jon. Automatic negation detection in narrative pathology reports. **Artificial Intelligence in Medicine**, v. 64, n. 1, p. 41-50, 2015.

PAMPOUCHIDOU, Anastasia et al. Depression Assessment by Fusing High and Low Level Features from Audio, Video, and Text. In **Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge**, p. 27-34, 2016.

PEDRINI, Hélio; SCHWARTZ, William R. **Análise de Imagens Digitais: Princípio, Algoritmos e Aplicações**. 1. ed. São Paulo, SP: Thomson Learning, 2008. 508 p.

PÉREZ, Jorge; ARENAS, Marcelo; GUTIERREZ, Claudio. Semantics and Complexity of SPARQL. In **International Semantic Web Conference**, p. 30-43, 2006.

PIANYKH, Oleg S. **Digital Imaging and Communications in Medicine (DICOM)**. Berlin: Springer, 2008. 383 p.

PONS, Ewoud et al. Natural language processing in radiology: a systematic review. **Radiology**, v. 279, n. 2, p. 329-343, 2016.

PRODANOV, Cleber C.; FREITAS, Ernani C. de. **Metodologia do trabalho científico: Métodos e Técnicas da Pesquisa e do Trabalho Acadêmico**. 2. ed. Novo Hamburgo, RS: Feevale, 2013. 276 p.

PUPPALA, Mamta et al. METEOR: an enterprise health informatics environment to support evidence-based medicine. **IEEE Transactions on Biomedical Engineering**, v. 62, n. 12, p. 2776-2786, 2015.

REIS, Adriana N. et al. Ferramenta para apoio ao diagnóstico baseada em processamento de imagens de tomografia computadorizada do tórax. In: IV ERCAS, 2016, Campo Bom, RS. **IV Escola Regional de Computação Aplicada à Saúde: Anais de resumos**, p. 20-21, 2016.

RIEGLER, Michael et al. Multimedia and Medicine: Teammates for Better Disease Detection and Survival. In: ACM Multimedia 2016, Amsterdam, Netherlands. **Proceedings of the 2016 ACM on Multimedia Conference**. New York, NY: ACM, 2016. p. 968-977.

ROEHRS, Alex et al. Personal Health Records: A Systematic Literature Review. **Journal of Medical Internet Research**, v. 19, n. 1, 2017.

RÖNNAU, Rodrigo F. **Segmentação Automática de Órgãos em Imagens de Tomografia Computadorizada do Tórax**. 2015. 81 f. Trabalho de Conclusão de Curso (Monografia) – Curso de Ciência da Computação, Universidade Feevale, Novo Hamburgo, RS, 2015.

RUBIN, Daniel L. et al. Medical Imaging on the Semantic Web: Annotation and Image Markup. In: **AAAI Spring Symposium: Semantic Scientific Knowledge Integration**, p. 93-98, 2007.

RUSSELL, Stuart; NORVIG, Peter. **Inteligência Artificial**. 3. ed. Rio de Janeiro, RJ: Elsevier, 2013. 988 p.

SHAUKAT, Furqan et al. Fully automatic detection of lung nodules in CT images using a hybrid feature set. **Medical Physics**, v. 44, n. 7, 2017.

SHOEMAKER, Allen L. What's Normal? Temperature, Gender, and Heart Rate. **Journal of Statistics Education**, v. 4, n. 2, 1996.

SNOMED. **What is SNOMED CT?**. 2017a. Disponível em: <<http://www.snomed.org/snomed-ct/what-is-snomed-ct>>. Acesso em: 09 abr. 2017.

SNOMED. **SNOMED CT Starter Guide**. 2017b. Disponível em: <<https://confluence.ihtsdotools.org/display/DOCSTART/SNOMED+CT+Starter+Guide>>. Acesso em: 09 abr. 2017.

SUND - LEVANDER, Märtha; FORSBERG, Christina; WAHREN, Lis K. Normal oral, rectal, tympanic and axillary body temperature in adult men and women: a systematic literature review. **Scandinavian Journal of Caring Sciences**, v. 16, n. 2, p. 122-128, 2002.

SUZUKI, Kenji. A supervised lesion-enhancement filter by use of a massive-training artificial neural network (MTANN) in computer-aided diagnosis (CAD). **Physics in Medicine and Biology**, v. 54, n. 18, 2009.

SUZUKI, Kenji. A review of computer-aided diagnosis in thoracic and colonic imaging. **Quantitative Imaging in Medicine and Surgery**, v. 2, n. 3, p. 163-176, 2012.

SZLOSEK, Donald A.; FERRETT, Jonathan. Using Machine Learning and Natural Language Processing Algorithms to Automate the Evaluation of Clinical Decision Support in Electronic Medical Record Systems. **eGEMs**, v. 4, n. 3, 2016.

TOPAZ, Maxim et al. Automated identification of wound information in clinical notes of patients with heart diseases: Developing and validating a natural language processing application. **International Journal of Nursing Studies**, v. 64, p. 25-31, 2016.

VENKATESH, Viswanath; DAVIS, Fred D. A theoretical extension of the technology acceptance model: Four longitudinal field studies. **Management Science**, v. 46, n. 2, p. 186-204, 2000.

VERMA, Amit; KHANNA, Gayatri. A survey on digital image processing techniques for tumor detection. **Indian Journal of Science and Technology**, v. 9, n. 14, 2016.

VIZZA, Patrizia et al. GIDAC: A prototype for bioimages annotation and clinical data integration. In: IEEE BIBM 2016, Shenzhen, China. **2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM16)**, 2016. p. 1028-1031.

WANG, Ying; YAO, Zhihong; LIU, Lei. Integration of DICOM and openEHR Standards. In: BOONN, William W.; LIU, Brent J. (Org.). **Medical Imaging 2011: Advanced PACS-based Imaging Informatics and Therapeutic Applications**. Orlando, FL: SPIE, 2011.

WAZLAWICK, Raul S. **Metodologia de Pesquisa para Ciência da Computação**. Rio de Janeiro, RJ: Elsevier, 2009. 124 p.

WEESE, Jürgen; LORENZ, Cristian. Four challenges in medical image analysis from an industrial perspective. **Medical Image Analysis**, v. 33, p. 44-49, 2016.

WEN, Hongwei et al. Multi-modal multiple kernel learning for accurate identification of Tourette syndrome children. **Pattern Recognition**, v. 63, p. 601-611, 2017.

WILLETT, Peter. The Porter stemming algorithm: then and now. **Program**, v. 40, n. 3, p. 219-223, 2006.

WILLIAMS, Mitch et al. A modular framework for clinical decision support systems: medical device plug-and-play is critical. **ACM SIGBED Review**, v.6, n. 2, 2009.

ZHAO, Jing et al. Learning from heterogeneous temporal data in electronic health records. **Journal of Biomedical Informatics**, v. 65, p. 105-119, 2017.