Image and Ontological Information Fusion for Cataract Surgery Recommendation

José Nuno Galveia
Centro Cirúrgico de Coimbra
Coimbra, Portugal

Luís A. da Silva Cruz
Instituto de Telecomunicações
Department of Electrical and Computer Engineering
University of Coimbra
Coimbra, Portugal, lcruz@deec.uc.pt

António Travassos
Centro Cirúrgico de Coimbra
Coimbra, Portugal

***Abstract*—Widely available digital ophthalmology data can be used to implement accurate Computer-Aided Diagnosis Systems. In this article we describe an automatic system which combines text clinical annotations, demographical information, as well as different types of ophthalmology image data to issue a recommendation for cataract surgery. Textual annotations are encoded using a standardized medical ontology nomenclature to enable higher level modeling. Image data is processed by convolutional neural networks to extract compact features. These two types of data together with demographical information are then inputted into a random forest classifier which then decides if surgery is recommended. The method proposed is evaluated on a real-life dataset, achieving accuracies and precisions around 90%. Several conclusions are drawn concerning the usefulness of the different input data types, used independently or combined.**

***Index Terms*—Ophthalmology, Information Fusion, Multimodal Image, Ontology, Cataract Surgery**

I. Introduction

In daily clinical practice medical ophthalmologists take into consideration patient current complaints, imaging data and past medical history to decide on which treatment options should be pursued, as illustrated in Fig. 1. We examine the question of whether clinical annotations expressed in a structured way and using ontologies can be combined with demographical data and image data to build a reliable cataract surgery recommendation system, following the approach outlined in Fig. 1. This work can be considered a case-study on the use of multimodal clinical information to construct computer-aided-diagnosis (CAD) systems. In the following sections we provide the details of the proposed system and present the results obtained evaluating its performance on a real-life dataset.

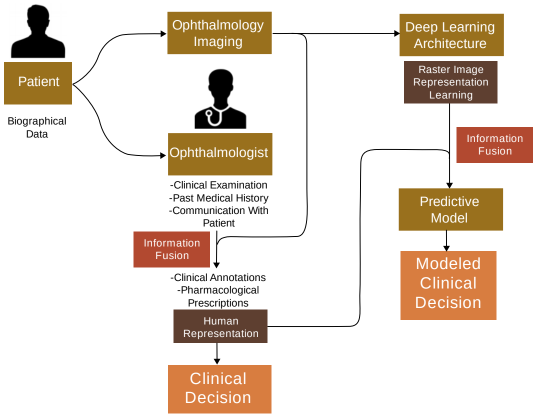

Fig. 1. Overview of clinical workflow and proposed decision model.

A. Related Work

Multiple systems have been developed to support clinical decision in ophthalmology. Medical evaluation combining multiple image modalities as input data is becoming more relevant as multimodal datasets become available [1]. Miri et al. [2] used information derived from optical coherence tomography (OCT) and color fundus photograph to segment the optic disc region. Suzuki et al. [3] used information from both OCT and infrared Scanner Laser Ophthalmoscopy (SLO) to classify pseudodrusen sub-types. Balaratnasingam et al. [4] have improved the definition of clinical phenotypes of cuticular drusen by combining different types of images, from fundus photography to electron microscopy. Ontology models and metrics have been applied to medical health record processing in [5] and Chan et al. in [6] used an ontology vector model based on the Systematized Nomenclature of Medicine Clinical Terms (SNOMED-CT) to improve the performance of clinical information indexing. Plastiras et al. [7] used an ontology based model to combine personal health records and electronic health records (EHR) for connectivity and interoperability. Several works fused clinical information and image data like Qi et al. [8] who combined three modalities of magnetic resonance imaging (MRI) and working memory clinical measures to obtain markers for working memory deficits in schizophrenia.

B. Objectives and Novelty of the Work

To the best of our knowledge integration of all commonly available patient data including biographical, clinical and image data into a single clinical predictive model has been used before only in our own work [9], which is here further developed in the context of a different application. Besides combining multimodal data, our approach proposes to convert

This work was partially funded by FCT project UID/EEA/50008/2019.

978-9-0827-9703-9/19/$31.00 ©2019 IEEE

all input data modalities into more compact representations, to be used in well performing classifiers like random forests. The relative importance of the different data types as well as the value added by their fusion to the solution of the problem at hand will also be researched. More concretely we address the following questions: Can multimodal ophthalmic image data be combined with medical records data to achieve higher accuracy in predicting the need for cataract surgery ? Is all information equally relevant for automated treatment recommendation ? Can data selection improve prediction performance ? Can representation learning be used to represent image information using compact features ? We answer these questions by designing, building and testing the automatic cataract surgery recommendation system described in the next sections.

II. Methodology

A. Model Description

We propose to build the system outlined in Fig. 2 which takes as inputs the three major data types identified before, *Demographical*, *Clinical Annotations* and *Image* and outputs a treatment recommendation decision. Fig.3 shows the processing chain in more detailed form, identifying the pre-processing needed to extract more compact data representations as well as the fusion and classification steps. It can be seen that the distinct input information types described in table I are processed by separate pipelines. Structured data is directly encoded in vector form. Unstructured clinical annotations are expressed using ontologies as described in Subsection II-B and multimodal image data are processed by several convolutional neural network (CNN) models (one for each image type) to compute compact features as described in Subsection II-C. Feature fusion is done by stacking data from selected input features. A forest of 200 randomized trees was chosen as the classifier for the final prediction. This choice was based on the good performance of this type of classifier, its low computational requirements and the intelligibility of the decisions.

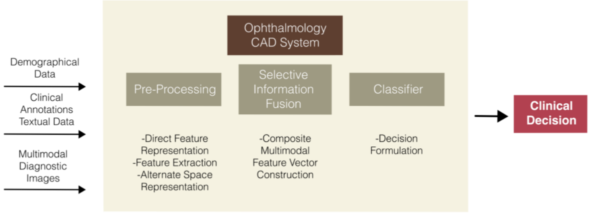
Demographical Data
Clinical Annotations Textual Data
Multimodal Diagnostic Images
Ophthalmology CAD System
Pre-Processing
Selective Information Fusion
Classifier
-Direct Feature Representation
-Feature Extraction
-Alternate Space Representation
-Composite Multimodal Feature Vector Construction
-Decision Formulation
Clinical Decision

Fig. 2. Overview of the proposed solution

B. Ontological Features

Medical annotations expressed in text in Portuguese were first mapped into standardized medical terms and then translated into SNOMED-CT concepts [10] and encoded as vectors containing a binary representation of the presence or absence of any element of the set of SNOMED-CT term codes. An abridged version of the SNOMED-CT ontology including the modeled cataract surgery procedure is presented in Fig. 4.

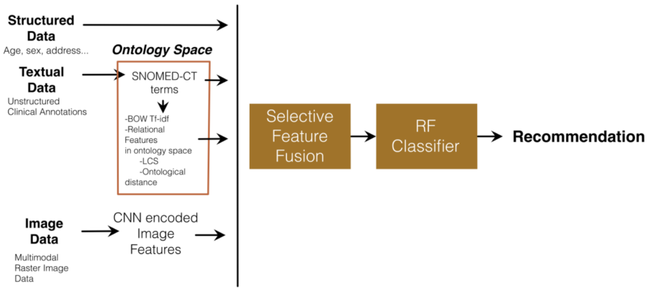

Fig. 3. Proposed Model; SNOMED-CT – Systematized Nomenclature of Medicine - Clinical Terms, BOW – Bag of Words, Tf-idf – Term Frequency Inverse Document Frequency, LCS – Least Common Subsumer, CNN – Convolutional Neural Network, RF – Random Forest.

TABLE I
Input Data Types

Data Type	Description
Structured	Demographic:Age, Sex, Civil State, Address
	Prescriptions: Previous pharmacological prescriptions by active principle
Unstructured	Annotations by Ophthalmologist: Clinical diary written in unformatted text describing clinical findings, conclusions and procedure recommendations.
Ontological	Ontological Information: Extracted from the **Unstructured** data including clinical concepts (diagnosis, findings and therapy) and encoded in SNOMED-CT accompanied by ontological distance from the procedure as well as the least common subsumer (LCS) computed between each concept and the labelled procedure.
Image	Ophthalmic Images: OCT, Slit lamp color photographs of the anterior segment of the eye, Scanning laser ophthalmoscopy (SLO) images of the retinal posterior pole in three bands: red-free blue reflectance (488 nm with low pass filter at 500 nm), autofluorescence (488 nm without filter) and infra-red (820 nm))

The procedure term code was used as reference for the computation of two ontological meta-features:

- Least common subsumer (LCS) [11] between each concept in the ontology vector and the modeled procedure concept – i.e. the most specific concept in the ontology hierarchy which is an ancestor of both concepts.
- Shortest distance (by counting edges in the ontology graph) between each concept in the ontology vector and the procedure concept.

C. Image Features

Five different raster image modalities listed in table I are used. Right and left eye images of the same modality for each observation were vertically stacked an resized to 512×256 pixels. A convolutional neural network (CNN) was trained separately for each raster image modality using labeled examples and used to output a 128 feature vector containing the activation values of the neurons in the third from last layer. The image feature extraction architecture is presented in Fig. 5. Network architecture parameters are listed in table II.

A CNN was chosen for its ability to implement an automated end-to-end representational learning of the image information. An alternative feature processing system based on autoencoders was tried without much success.

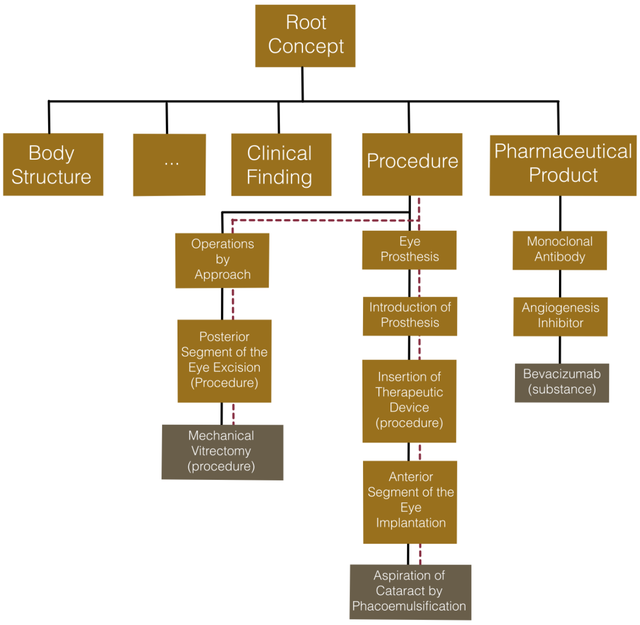
Root Concept
Body Structure
...
Clinical Finding
Procedure
Pharmaceutical Product
Operations by Approach
Eye Prosthesis
Monoclonal Antibody
Introduction of Prosthesis
Angiogenesis Inhibitor
Posterior Segment of the Eye Excision (Procedure)
Insertion of Therapeutic Device (procedure)
Bevacizumab (substance)
Mechanical Vitrectomy (procedure)
Anterior Segment of the Eye Implantation
Aspiration of Cataract by Phacoemulsification

Fig. 4. Abridged example diagram of the SNOMED-CT Ontology including the modeled cataract procedure and a hypothetical vitrectomy procedure; dashed line represents the path from the procedures to their LCS – Procedure; shortest distance between the aforementioned concepts can be computed by counting the number of graph edges along the dashed path

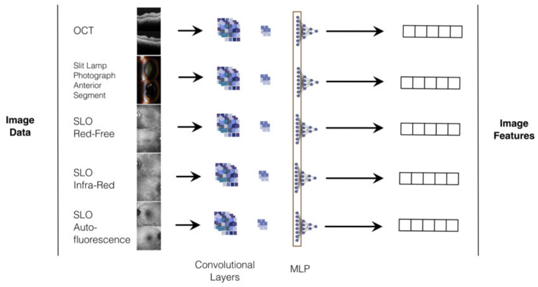

Fig. 5. CNN image feature extractor; OCT – Optical Coherence Tomography, SLO – Scanning Laser Ophthalmoscopy, MLP – Multi-layer Perceptron

TABLE II
TRAINED CNN ARCHITECTURE. LAYER TYPE: I-INPUT, C-CONVOLUTIONAL, MP-MAX-POOLING, D-DROPOUT (RATE=0.3), FC-FULLY CONNECTED

Layer	Type	Maps and Neurons	Filter Size
0	I	1 M x 512 x 256 N	-
1	C	32 M x 15 x 15 N	15 x 15
2	MP	32 M x 8 x 8 N	8 x 8
3	C	64 M x 5 x 5 N	5 x 5
4	D	64 M x 5 x 5 N	5 x 5
5	MP	64 M x 4 x 4 N	4 x 4
6	C	128 M x 3 x 3 N	3 x 3
7	D	128 M x 3 x 3 N	3 x 3
8	MP	128 M x 3 x 3 N	3 x 3
9	C	128 M x 2 x 2 N	2 x 2
10	FC	128 N	1 x 1
11	FC	64 N	1 x 1
12	FC	2 N	1 x 1

D. Training and Testing Dataset

TABLE III
MODEL PERFORMANCE WITH DIFFERENT INPUT DATA; RESULTS ARE THE AVERAGE OF TEN MODEL RUNS; OCT – OPTICAL COHERENCE TOMOGRAPHY, AS – ANTERIOR SEGMENT SLIT LAMP PHOTOGRAPH, RF – RED-FREE SLO IMAGE, IR – INFRA-RED SLO IMAGE, AF – AUTOFLUORESCENCE SLO IMAGE

Classifier Input Data	Acc. (%)	Prec. (%)	Recall (%)	F1 (%)
Structured Data	64.65	82.01	64.08	71.94
Ontological Data	83.88	89.34	82.84	85.97
Ontological Meta Data	83.88	89.34	82.84	85.97
All Ontological Data	83.56	89.37	82.39	85.73
All Non Image Data	**86.04**	**91.59**	**84.46**	**87.88**
CNN Features OCT	78.46	82.91	79.12	80.97
CNN Features AS	84.7	**86.52**	85.89	86.2
CNN Features RF	84.74	85.85	86.45	86.15
CNN Features IR	81.17	82.51	83.27	82.89
CNN Features AF	78.17	79.32	80.82	80.06
All CNN Features	**86.43**	86.31	**88.82**	**87.55**
Non Image + Learned	88.46	88.59	90.35	89.46
All Non Medical Features	86.88	86.67	89.28	87.95

The data used to build the prediction model consisted of 17,470 unique EHR of patients with average age of 69.73±14.31 and a male to female ratio of 46.14%/53.86%. Records were divided into two equal parts representing patients who did and who didn't undergo the procedure after medical observation. Information of the data types listed in table I were recorded for each observation.

E. Training and Computation Times

A split using 80:20 ratio into training and test subsets was used for cross-validation in every experiment. Network training and classification operations were performend using the Keras framework and Tensorflow as the backend processor. All computations were done using a computer equipped with an Intel core i7-6700 CPU and an Nvidia 1080 GPU. The computation time for all ontological features 0.14 ms per record. The image representation learning using CNNs took an average of 25 minutes per epoch. All models were trained for 50 epochs or until no further improvement in model accuracy was detected. After training, feature computation time per image was 1.9 ms. Training the random forest used in the final classification task took no more than 4.86 s for worst case scenario (i.e. including all possible features). After training, in the worst case scenario (i.e. including all features), the classifier took an average of 0.14 s per patient to compute the final prediction.

III. RESULTS

In a first exploration we examined the performance of all models including all available input features as well as all possible features in each considered feature subset: structured data, ontological data, ontological meta data, each individual image modality and all image derived features. Ten runs of the training and testing steps were performed and the corresponding test-time classification/recommendation performance indicators computed. The average results are presented in table III using the usual classification performance indicators *Accuracy*, *Precision*, *Recall* and *F1 Score*.

Ontological data and meta-data enable an accuracy that is in the same performance tier as image derived features. It can be observed that ontological information compression in the

TABLE IV

Top-ten features in decreasing order of importance; VA – visual acuity, OCT – optical coherence tomography, AS – anterior segment slit lamp photograph, RF – red-free SLO image, IR – infra-red SLO image, AF – autofluorescence SLO image; the numbers after the terms OCT, RF, AF and IR represent the index of a given feature in the 128 feature vector generated by representation learning

Feature Subset		
Global	**Image**	**Ontological**
1 - Ofloxacin	1 - RF 104	1 - Ofloxacin
2 - VA testing	2 - RF 19	2 - VA testing
3 - RF 104	3 - RF 95	3 - Tonometry/Ofloxacin
4 - RF 19	4 - RF 98	4 - Pseudophakia
5 - Tonometry Ofloxacin	5 - AS 21	5 - Diabetes Mellitus II
6 - RF 95	6 - AS 91	6 - Moxifloxacin
7 - RF 98	7 - OCT 109	7 - Photocoagulation of the Retina
8 - AS 21	8 - RF 49	8 - Mechanical Vitrectomy
9 - AS 91	9 - RF 113	9 - Levofloxacin
10 - OCT 109	10 - IR 42	10 - Posterior Segment Fluorescein Angiography

TABLE V

Maximum Model Performance Values; values in parenthesis represent the number of features incorporated in the model that reached the specified performance

Feature Subset	Performance	
All (n=1020)	Accuracy (%)	89 (504)
	F1 Score (%)	90 (504)
	Precision (%)	90 (504)
	Recall (%)	92 (720)
Non medical (n=644)	Accuracy (%)	88 (278)
	F1 Score (%)	89 (452)
	Precision (%)	93 (2)
	Recall (%)	90 (379)
Ontological (n=124)	Accuracy (%)	82 (113)
	F1 Score (%)	84 (113)
	Precision (%)	86 (107)
	Recall (%)	88 (2)
Image (n=640)	Accuracy (%)	87 (174)
	F1 Score (%)	88 (542)
	Precision (%)	93 (2)
	Recall (%)	90 (354)

form of ontological meta-features does not impact any of the models performance indices, yielding exactly the same average performance. Concerning precision, surprisingly, features derived from most image modalities lead to slightly inferior results than those obtained using either ontological data or all non-image data. This effect was not present for the case of the accuracies, where image-related information provided better results than non-image data. It is well known that a previous feature selection should be done when dealing with high dimensional datasets. To rank features by classification usefulness, for each subset we trained a decision tree and rated each input feature's importance according to its depth in the constructed decision graph [12]. The top ten features for each subset are presented in table IV ordered by decreasing usefulness.

Building upon the previous experiment we decided to research the influence of feature partition into the several subsets described before on the maximum attainable classification/recommendation performance. Several classifiers were trained using first all data features, then only the non medical features (image and demographical data), the ontological features and also only image-related features. In all cases we recorded the number of features used in the classifier that reached the highest performance. The results of this experiment are presented in table V organized per feature subset where in the second column the performance indicators are listed accompanied by the number of features for the maximum performance classifier.

It can be concluded that a relatively large number of the entire set of input features are needed for maximum model accuracy. For instance 504 of all the global features are needed to reach an accuracy of 0.89. For the non medical and image subsets only 2 features are needed to achieve a model precision of 0.93 for both instances. It is also noticeable that the proposed model can reach a maximum accuracy of 0.89 using a subset of features chosen from all available clinical information and 0.88 on a subset of features constructed without medical information input. These numbers show that combining features does not always lead to better performance indicators and that non medical data alone can provide performance similar to that obtained based on the image data.

IV. Discussion

Computer aided diagnostic and recommendation systems leveraging the different types of input information in an integrated fashion will become an increasingly important asset in daily clinical practice. Such systems will be able to integrate clinical information at a scale beyond the abilities of any clinician and after careful validation they will provide an objective clinical opinion. Clinical validation presupposes that the inner workings of a proposed CAD system must be amenable to scrutiny by clinicians not only to attest its validity but also to troubleshoot possible failures. We explored the possibilities afforded by a multimodal dataset in the construction of an integrated 'white box' modeling algorithm for clinical event prediction in ophthalmology. Preprocessing allowed us to represent input data with compact and discriminative feature vectors amenable to fusion. Models built from different feature subsets gave us the possibility to probe the significance of different input features. Medical features were in general the most important for model accuracy. Ontological meta-features enabled a model performance that closely matched the one afforded by non processed ontological features. This shows that hierarchical feature representation in ontological space can preserve discriminating ability with regards to the original input data. We expected anterior segment photographs to be the most informative for cataract diagnosis but in our model posterior segment RF SLO imaging derived features carried the higher discriminative ability. Our interpretation is twofold. On one hand slit lamp photograph is not as standardized making the construction of a compact discriminative feature set more challenging. On the other hand SLO RF imaging of the eye fundus can provide a more standardized image set that can indirectly provide information concerning the optical attenuation properties of the lens.

With regards to ontological features, since in clinical practice all invasive ophthalmic procedures are preceded by the topical administration of an antibiotic (in the this case ofloxacin) which is recorded in the clinical annotations, the association between ofloxacin-related features and procedure recommendation was expected. Visual acuity testing is also a procedure that can carry significant diagnostic and prognostic implications for procedure recommendations and as such its high predictive impact was expected. It is common knowledge among ophthalmologists that diabetes accelerates the progression of cataract disease, that retinal photocoagulation is a procedure performed in patients with advanced diabetic retinopathy and that fluorescein angiography is a critical exam in the evaluation of advanced diabetic retinopathy. These observations show that the decision model operation reflects the feature importance reported in Table IV and is amenable to rational explanation and in line with common clinical knowledge. With reference to feature selection for model performance optimization two major categories of interpretations can be drawn from the data: feature subset influence in global performance and number of features required for maximum model performance – information compression and relevance. It was concluded that a relatively large amount of input features is needed to ensure maximum model accuracy. While clinical diagnosis of cataract is straightforward from anterior segment observation at the slit lamp, a pondered recommendation taking into consideration the potential visual impact of the disease and the prognosis of an eventual surgical intervention can only be made with access to information of the posterior segment. In all cases a careful evaluation of all considered image modalities is required for adequate surgical planning and prognosis implications. These observations can be relevant for data management in clinical environments as they show than not all data is equally important for a given clinical decision. Relevant image modalities can be stored in higher resolution lossless formats and made immediately available for consultation while less relevant formats can be available in compressed form or represented only by descriptive feature. Also if careful feature selection is implemented, considering specific needs, computational requirements can be lighter, allowing portable implementations for use in remote areas.

V. Conclusion

Selective combination of multimodal EHR data in an integrated model can provide good accuracies in the recommendation of cataract surgery, one of the most commonly performed ophthalmic procedures. Our study reveals that clinical annotations in the form of ontological encodings are the most relevant features regarding clinical event prediction accuracy. Furthermore we show that ontological encodings of clinical annotations and ontological meta-features are effective in the representation of medical knowledge. The experiments show that a convolutional deep learning architecture can be used to extract sparse representations of input image data in an automated end-to-end approach enabling practical effective multimodal image fusion. It is also demonstrated that the proposed solution can be fine tuned by exploring the relative impact of each feature subset in the model's performance. We also showed that the use of random forests provides a way to implement a recommendation system which is not opaque enabling some understanding of the processing steps that lead to final decision formulation. This enables auditability and internal understanding by experienced clinicians. Taking into consideration practical usefulness, performance, computational requirements, computation times, modularity and auditability of the proposed system we believe that it can be used for real time CAD and decision support in ophthalmology.

VI. Compliance with Ethical Standards

The authors have no conflicts of interest to declare. Clinical records were adequately anonymized and processed retrospectively with no impact in the standard of care. This study was approved by the *Centro Cirúrgico de Coimbra* institutional review board.

References

[1] A. P. James and B. V. Dasarathy, "Medical image fusion: A survey of the state of the art," *Information Fusion*, vol. 19, pp. 4–19, 2014.

[2] M. S. Miri, M. D. Abràmoff, K. Lee, M. Niemeijer, J. K. Wang, Y. H. Kwon, and M. K. Garvin, "Multimodal Segmentation of Optic Disc and Cup from SD-OCT and Color Fundus Photographs Using a Machine-Learning Graph-Based Approach," *IEEE Transactions on Medical Imaging*, vol. 34, no. 9, pp. 1854–1866, 2015.

[3] M. Suzuki, T. Sato, and R. F. Spaide, "Pseudodrusen subtypes as delineated by multimodal imaging of the fundus," *American Journal of Ophthalmology*, vol. 157, no. 5, pp. 1005–1012, 2014.

[4] C. Balaratnasingam, S. Cherepanoff, R. Dolz-Marco, M. Killingsworth, F. K. Chen, R. Mendis, S. Mrejen, L. K. Too, O. Gal-Or, C. A. Curcio, K. B. Freund, and L. A. Yannuzzi, "Cuticular Drusen: Clinical Phenotypes and Natural History Defined Using Multimodal Imaging," *Ophthalmology*, vol. 125, no. 1, pp. 100–118, 2018.

[5] M. Gurcan, B. Smith, S. Arabandi, M. Brochhausen, M. Calhoun, P. Ciccarese, S. Doyle, B. Gibaud, I. Goldberg, C. Kahn, J. Overton, and J. Tomaszewski, "Biomedical imaging ontologies: A survey and proposal for future work," *Journal of Pathology Informatics*, vol. 6, no. 1, p. 37, 2015.

[6] L. W. Chan, Y. Liu, C. R. Shyu, and I. F. Benzie, "A SNOMED supported ontological vector model for subclinical disorder detection using EHR similarity," *Engineering Applications of Artificial Intelligence*, vol. 24, no. 8, pp. 1398–1409, 2011.

[7] P. Plastiras and D. M. O'Sullivan, "Combining Ontologies and Open Standards to Derive a Middle Layer Information Model for Interoperability of Personal and Electronic Health Records," *Journal of Medical Systems*, vol. 41, no. 12, pp. 1–15, 2017.

[8] S. Qi, V. D. Calhoun, T. G. Van Erp, J. Bustillo, E. Damaraju, J. A. Turner, Y. Du, J. Yang, J. Chen, Q. Yu, D. H. Mathalon, J. M. Ford, J. Voyvodic, B. A. Mueller, A. Belger, S. McEwen, S. G. Potkin, A. Preda, T. Jiang, and J. Sui, "Multimodal Fusion with Reference: Searching for Joint Neuromarkers of Working Memory Deficits in Schizophrenia," *IEEE Transactions on Medical Imaging*, vol. 37, no. 1, pp. 93–105, 2018.

[9] J. N. Galveia, A. Travassos, and L. A. da Silva Cruz, "An ophthalmology clinical decision support system based on clinical annotations, ontologies and images," in *2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS)*, June 2018, pp. 94–99.

[10] T. Pedersen, S. V. S. Pakhomov, S. Patwardhan, and C. G. Chute, "Measures of semantic similarity and relatedness in the biomedical domain," *Journal of Biomedical Informatics*, vol. 40, no. 3, pp. 288–299, 2007.

[11] F. Baader, B. Sertkaya, and A. Y. Turhan, "Computing the least common subsumer w.r.t. a background terminology," *Journal of Applied Logic*, vol. 5, no. 3, pp. 392–420, 2007.

[12] M. Dash and H. Liu, "Feature selection for classification," *Intelligent Data Analysis*, vol. 1, pp. 131–156, 1997.