Comprehensive Framework for the Fusion of Clinical Data

Anam Haq

Home University
Poznan University of Technology (PUT)

Szymon Wilk

Host University
Polytechnic University of Catalonia (UPC)

Alberto Albelló

Outline

- Introduction
- Problem Statement
- Goal
- Literature Review
- General Fusion Framework
- Identified Challenges
- Industrial Project
- Developed Case Studies

Introduction: Data Fusion

“Integration of data and knowledge from multiple sources of diversified format and structure”

- Data fusion and information fusion often used as synonyms but they are different
  - a. Data fusion typically used for raw data that need *pre-processing* before integration.
  - b. Information fusion used for already processed data

- We will be handling *data fusion* related to *medical domain*

Problem Statement

The singular modality in diagnostic clinical system is insufficient and does not correspond to accurate prediction of clinical diseases

Challenges Identified:

1. Heterogeneity and Complexity
2. Imbalanced Class Structure
3. Data incompleteness

Research Goals

To improve the performance of computational Health care standards in medical diagnostic systems such as monitoring, identification and classification.

Challenges in Clinical Data Fusion

Overwhelming presence and yet *underutilization* of clinical data

1. Frequently data is imbalance
2. Data is imperfect (inconsistent and incomplete)
3. Significant heterogeneity and complexity of data stored in clinical information systems (CIS):
    - *Structured clinical records*
    - *Unstructured clinical notes*
    - *Medical images*

Literature Review

- Traditional approaches to data fusion [3]
  1. Combination of data (COD)
     - Aggregation of Features into a common vector.
  2. Combination of interpretation (COI)
     - A separate classifier is constructed for each considered data source.
     - Outputs are aggregated by a combiner.
- Drawbacks [4,5]:
  1. Curse of Dimensionality (Combination of Data)
  2. Inability to handle inter-source dependencies (Combination of Interpretation)

General Fusion Framework

Bring every data source to a common knowledge representation by using diversified transformations

- Steps in General Fusion Framework [4]:
  1. Transformation of data into a *meta-space*
  2. Fusion of the meta-space *(meta-space fusion)*
  3. Construction of a *meta-classifier*

General Fusion Framework

1 Meta Fusion

- Transforming data from heterogeneous space to a homogenous space where it can be integrated

2 Meta Space Fusion

- Integrating features present in a homogenous space

3 Meta Classification

- Construction of one or more classifiers using the fused feature vector

Limitations of General Fusion Framework

1. Unable to identify which algorithm should be used at each step
2. Choice of data fusion method is not obvious as it highly depends on the properties of the data
3. Unable to handle issues with clinical data
   - Imbalance Data
   - Data Imperfection

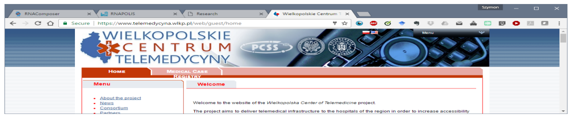
Objective

1. Evaluating the *performance* of currently developed data fusion techniques on clinical data
2. Developing a *control model* to handle problems associated with data fusion technique selection by improvising the general fusion framework
3. Developing a *comprehensive framework* for clinical data fusion addressing additional problems and challenges:
   - Presence of imbalance classes
   - Noise
   - Missing values
   - Outliers

Developed Case Studies

1 Evaluation of the performance of data fusion using the data fusion approach

- Prediction of the ***type of treatment*** (surgical vs. non-surgical) for patients with bone fractures using a decision model derived from the ***fusion of image and non-image data***
- *Data provided by the Wielkopolska Center of Telemedicine (a tele-consultation platform for patients with multiple injuries), Poznan, Poland*



2 Extraction and selection of relevant features from clinical images

- Detection of ***wet aged-related macular degeneration*** using combination of ***structural and textural properties*** of retinal pigment epithelium layer
- *Data was obtained from Army Forces Institute of Ophthalmology, Rawalpindi, Pakistan*

Case Study: 1

Building a prediction model to detect the type of treatment a patient should undergo using the information obtained from image and non-image data

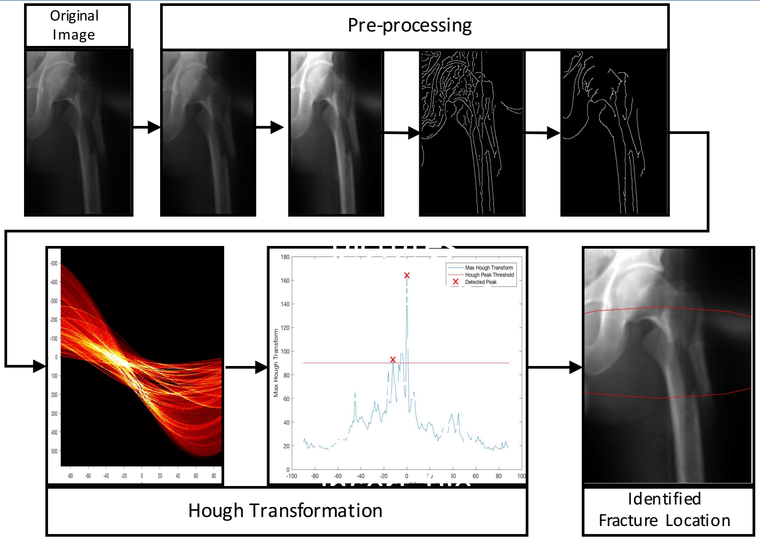
1 Extraction of feature from the X-Ray images using various image process transformations (Image Data)

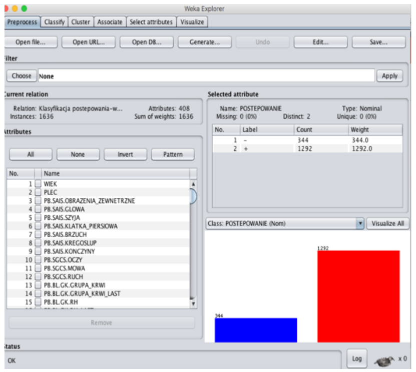
2 For clinical records following steps are considered:

- *Discretization* of numerical features using experts opinion
- Introducing *additional features* capturing information about injuries
- Removal of *"useless"* features (e.g., features with the majority of missing values)

3 Combining features obtained from images and clinical records using *combination of data* approach of data fusion

Results





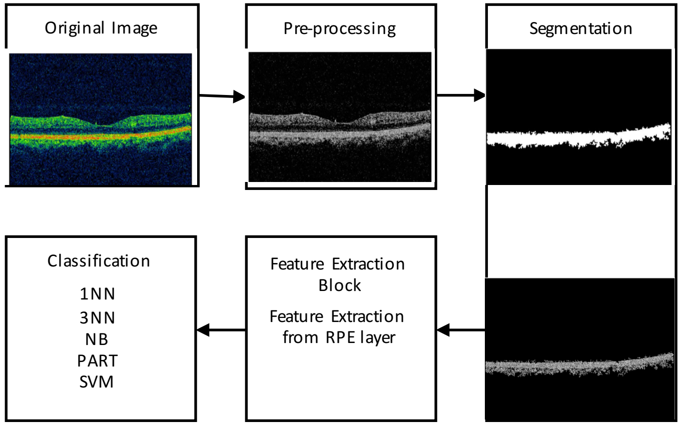
Classifier	Features	Overall Acc (%)	Non-Surgical (%)	Surgical (%)	G-Mean (%)
NB	Image	78.64	**77.50**	79.40	78.44
	Clinical	78.64	62.50	88.90	74.54
	Fused	**83.50**	67.50	**93.70**	**79.53**
SVM	Image	79.61	**80.00**	79.40	79.70
	Clinical	67.96	62.50	71.40	66.81
	Fused	**80.58**	77.50	**82.50**	**80.00**
DT	Image	78.64	**82.50**	76.20	**79.29**
	Clinical	66.20	60.00	69.80	64.71
	Fused	**79.61**	75.00	**82.50**	78.66
RF	Image	70.87	65.00	74.60	69.60
	Clinical	75.72	57.50	**87.30**	70.85
	Fused	**79.61**	**82.76**	78.40	**80.55**

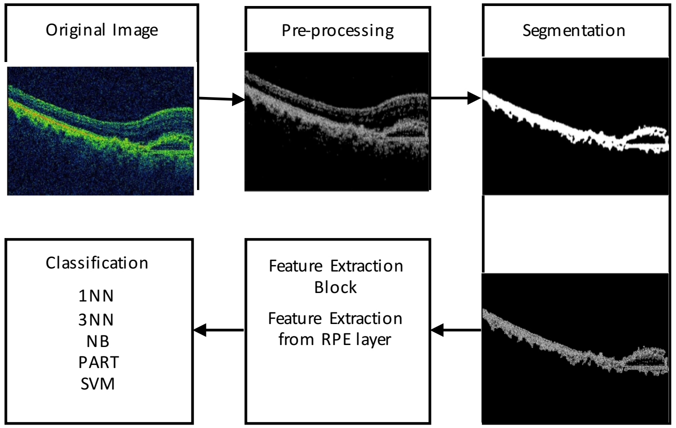
Case Study: 2

Building a diagnostic model to detect Wet Aged Related Macular Degeneration using Optical Coherence Tomographic Images

1. Extraction of *textural* and *structural* features from the identified abnormalities
2. Extraction of *textural* and *structural* features from Retinal Pigment Epithelium Layer
3. Building a classifications model for each type of features and their combinations

Results





Feature set	Classifier	Acc [%]	Sens [%]	Spec [%]	G-Mean
Abnormalities (S+T)	3-NN	84.3	84.2	84.4	84.3
RPE Layer (S)	3-NN	92.2	94.7	90.6	92.6
RPE Layer (S+T)	**3-NN**	**94.1**	**95.0**	**93.5**	**94.3**
Abnormalities (S+T)	NB	88.2	86.0	90.0	87.8
RPE Layer (S)	NB	94.1	95.0	93.0	94.3
RPE Layer (S+T)	**NB**	**96.1**	**91.3**	**100.0**	**95.5**
Abnormalities (S+T)	PART	88.2	85.7	90.0	87.8
RPE Layer (S)	PART	92.2	90.5	93.3	91.9
RPE Layer (S+T)	**PART**	**94.2**	**91.9**	**96.5**	**93.7**
Abnormalities (S+T)	SVM	86.3	88.9	84.8	86.8
RPE Layer (S)	SVM	78.4	75.0	80.6	77.8
RPE Layer (S+T)	**SVM**	**94.1**	**100.0**	**93.7**	**96.8**

Structural (S), Textural (T)

Industrial Project

1. Building diagnostic model for *detection of lung cancer* and predicting the patients *survival rate* using data fusion using *deep learning* approaches

2. Building a *diagnostic model* for the *detection of lung fibrosis*

Papers Accepted and Under-Review

1. Fusion of clinical data: A case study to predict the type of treatment of bone fractures (*Accepted*)
   - Int. Workshop on Data Science: Methodologies and Use-Cases (DaS) collocated with ADBIS 2017 (Communications in Computer and Information Science (CCIS), Springer)
2. Detection of wet age-related macular degeneration in OCT images: A case study (*Under-Review*)
   - Submitted to the Innovations in Bio-Medical Engineering (IiBE) 2017 conference (Advances in Intelligent and Soft Computing (AISC), Springer)

Timeline

Sessions	Details	Status
Fall [Oct (2016)-Feb (2017)]	Submission of DPP and literature review (done)	Completed
Spring [Mar(2017)-Jul(2017)]	Submission of a conference paper to DaS 2017 (ADBIS) (accepted) Submission of a conference paper to IiBE-2017 (under review)	Completed
Fall [Sept(2017)-Feb(2018)]	Submission of TPR (done) Moving to the host university (planned) Submission of a conference paper on evaluation and comparison of existing data fusion techniques (IEEE-HealthCom-18) Submission of a journal paper on a control model development for clinical data fusion (Artificial Intelligence in Medicine, Journal of Biomedical Informatics)	In progress
Spring [Mar(2018)-Jul(2018)]	Submission of a conference paper on the idea and initial evaluation of a comprehensive framework for clinical data fusion (IEEE BIBM-18)	Planned
Fall [Sept(2018)-Feb(2019)]	Submission of a journal paper on a comprehensive framework for fusion of clinical data (Artificial Intelligence in Medicine, Journal of Biomedical Informatics)	Planned
Spring [Mar(2019)-Oct(2019)]	Thesis write-up and defence	Planned

References

1. Castanedo, "A review of data fusion techniques," *The scientific World Journal,* 2013
2. D. Lahat, T. Adali and C. Jutten, "Multimodal Data Fusion: An Overview of Methods, Challenges, and Prospects," in *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1449-1477, Sept. 2015
3. T. Rohlfing, A. Pfefferbaum, E. Sullivan,C. Maurer, "Information fusion in biomedical image analysis: combination of data vs combination of interpretations," *19th International Conference on Information Processing in Medical Imaging (IPMI'05)*, 2005.
4. G. Lee, A. Madabhushi, "A knowledge representation framework for integration, classification of multi-scale imaging and non-imaging data: Preliminary results in predicting prostate cancer recurrence by fusing mass spectrometry and histology," *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 2009.
5. P.Tiwari, S. Viswanath, G. Lee, A. Madabhushi, "Multi-model data fusion schemes for integrated classification of imaging and non-imaging biomedical data," *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 2011
6. Rafael C. Gonzalez, Digital Image Processing (3rd Edition).
7. Anu T.c., Mallikarjunaswamy M.s. and Rajesh, “Detection of Bone Fracture using Image Processing Methods,” IJCA Proceedings on National Conference on Power Systems and Industrial Automation, NCPSIA 2015(3):6-9..
8. San myint, aung soe khaing, hla myo tun, “Detecting leg bone fracture in x-ray images”, International journal of scientific & technology research volume 5, issue 06, june 2016