



MACHINE LEARNING PROJECT

Team members

- Kasma Marianna
- Tsekouras Ilias
- Tsiolis Spyridon



ARGUMENTATION MINING

- Data Annotation
- Data pre-processing
- Classification
- Clustering
- Discussion/Comments

DATA ANNOTATION



Data Source: Horizon 2020

- ✓ Gathered abstracts originated from different scientific fields, concerning environmental, biomedical and other scientific researches.

Data Labeling Tool:



Label Studio

Structure:

Background^[1]

Objective/Aim^[2]

Method^[3]

Result^[4]

Conclusion^[5]

Argument:

Evidence^[6]

Claim^[7]

or 'Neither'

Abstract:

Models for cardiac mechanics require an activation mechanism properly representing the stress-strain relations in the contracting myocardium. **background**

In this paper, we propose a new activation model that accounts for the transmural heterogeneities observed in myocardial strain measurements. **objective**

In order to take the anisotropy of the active mechanics into account, our model is based on an active strain formulation. **method**

Thanks to multiplicative decomposition of the deformation gradient tensor, in this formulation, the active strains orthogonal to the fibers can be naturally described. **method**

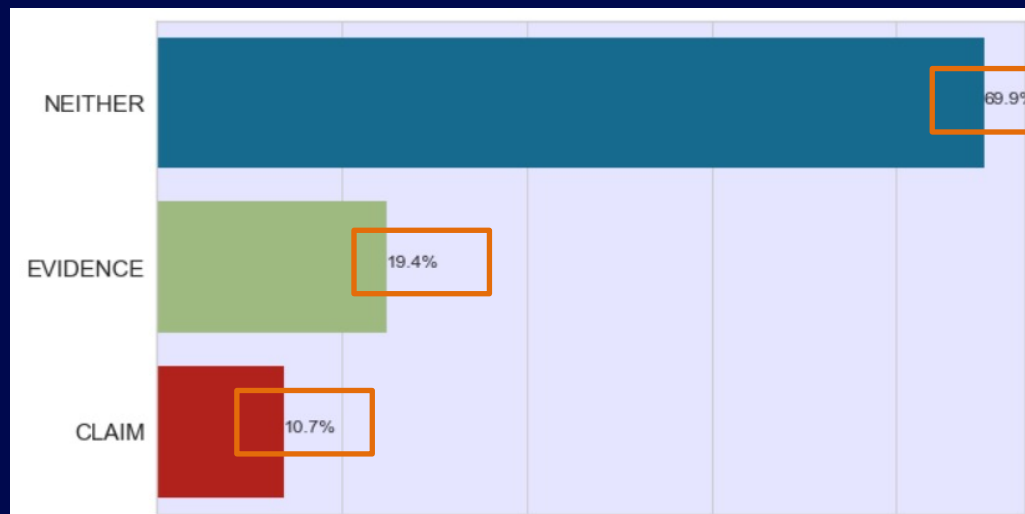
We compare the results of our novel formulation against different anisotropic models of the active contraction of the cardiac muscle, as well as against experimental data available in the literature. **method**

We show that with the currently available models, the strain distributions are not in agreement with the reported experimental measurements. **result**

Conversely, we show that our new transmurally heterogeneous orthotropic activation model improves the accuracy of shear strains related to in-plane rotations and torsion. **result**



DATA OVERVIEW



- 2686 Abstracts
- 32004 Sentences
- 22375 (Neither)
- 6210 (Evidence)
- 3419 (Claim)



DATA PRE-PROCESSING

Removed
punctuation
marks

Split
sentences
into tokens

Removed
stopwords
and digits

Lemmatized
verbs and
nouns

Split into
train-test
sets

One-hot
encoding of
the labels

Tokens into
sequence of
number

Padding
length
sequences



CLASSIFICATION

TWO CNN MODELS

1st : Custom Created Embeddings

2nd : Pre-Trained Embeddings
(GloVe)



MODEL SUMMARY

- Sequential Model
- Embedding Layer
- Dropout Layer
- Convolutional Layer
- GlobalMaxPooling
- Dense Layer



CLASSIFICATION

1st Model

Confusion Matrix

Label	Neither	Evidence	Claim
Neither	6094	491	128
Evidence	797	1018	48
Claim	594	182	250

Accuracy Score : 76.7%

2nd Model

Confusion Matrix

Label	Neither	Evidence	Claim
Neither	6304	284	125
Evidence	994	820	49
Claim	666	98	262

Accuracy Score : 76.9%



CLASSIFICATION

BASELINE MODEL

CREATION

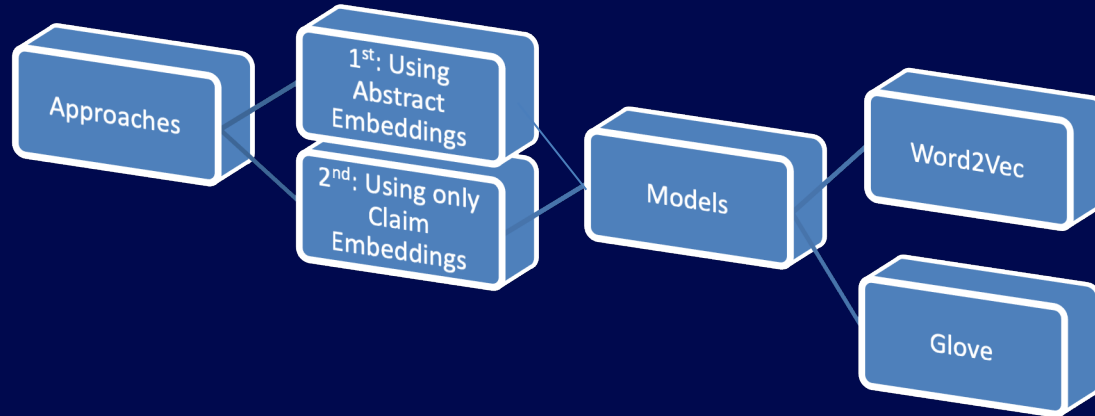
- Word Lexicons
 - Claim (provide, confirm, suggest etc.)
 - Evidence (results, findings, etc.)
- Label all last sentences as claim



RESULTS

	Recall	Precision	F- Score
Claim	0.594	0.470	0.525
Evidence	0.084	0.333	0.135

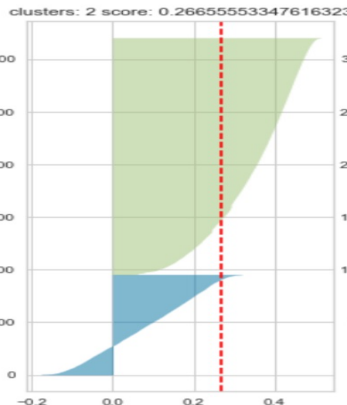
CLUSTERING



CLUSTERING

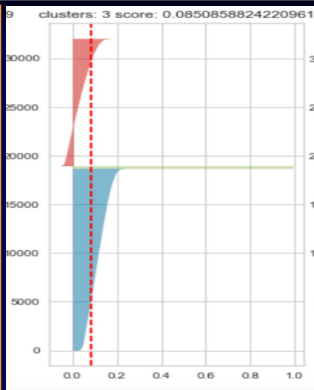
Abstract – Word2Vec

- $K = 2$
- Silhouette val = 0.26



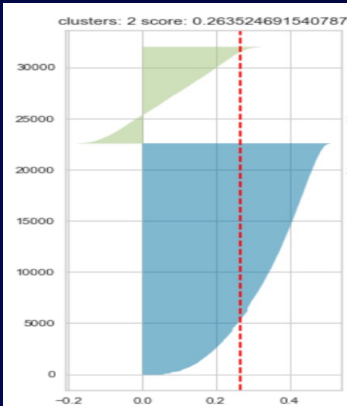
Abstract – GloVe

- $K = 3$
- Silhouette val = 0.08



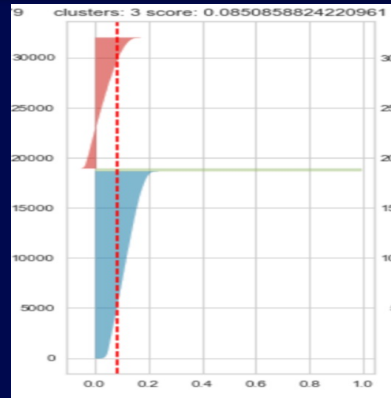
Claim – Word2Vec

- $K = 2$
- Silhouette val = 0.26



Claim – GloVe

- $K = 3$
- Silhouette val = 0.08





DISCUSSION/COMMENTS

- More attributes to be used in the future in our clustering
- Silhouette Values close to zero
- Compared to the dummy classifier we observed only 10% improvement in our models

Thank you for your attention!



Any Questions?