Multimodal Detection of Information Disorder from Social Media

Kirchknopf Armin
Institute of Creative Media Technologies
University of Applied Sciences
St. Pölten, Austria
armin.kirchknopf@fhstp.ac.at

Slijepčević Djordje
Institute of Creative Media Technologies
University of Applied Sciences
St. Pölten, Austria
djordje.slijepcevic@fhstp.ac.at

Zeppelzauer Matthias
Institute of Creative Media Technologies
University of Applied Sciences
St. Pölten, Austria
matthias.zeppelzauer@fhstp.ac.at

CBMI'21 (Content-Based Multimedia Indexing)

211012 Chia-Chun Ho

Outline

Introduction

Related Work

Proposed Approach

Experiments

Conclusion

Comments

Introduction

Fake news detection

- Like the U.S. presidential election in 2016 the public has become aware of impact that fake news have on public opinion.
- Due to the ever-increasing amount of data, automated analysis approaches are necessary to assist the detection and verification of fake news.
- In context of this paper, focus on fake news in terms of information disorder as defined by Wardle.

Introduction
Information disorder

- Three types of information disorder can be distinguished:
  - Misinformation
    - Refers to misleading content produced without a specific intent.
  - Disinformation
    - Refers to purposely generated and potentially harmful content.
  - Malinformation
    - Harmful content including hate speech and harassment.

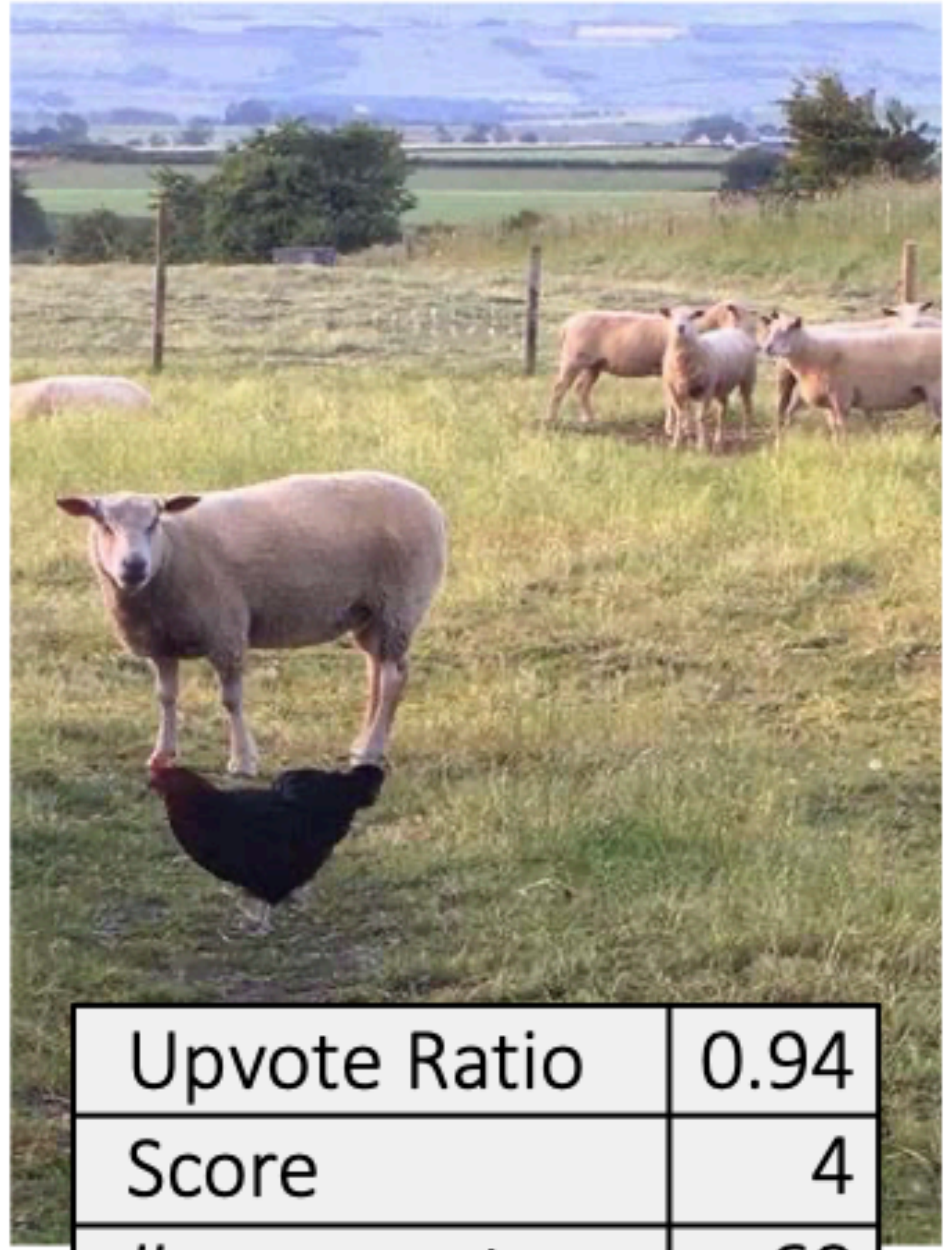
Introduction

Contribution

- An end-to-end learnable modular approach which combine multiple heterogeneous modalities for the detection of information disorder.

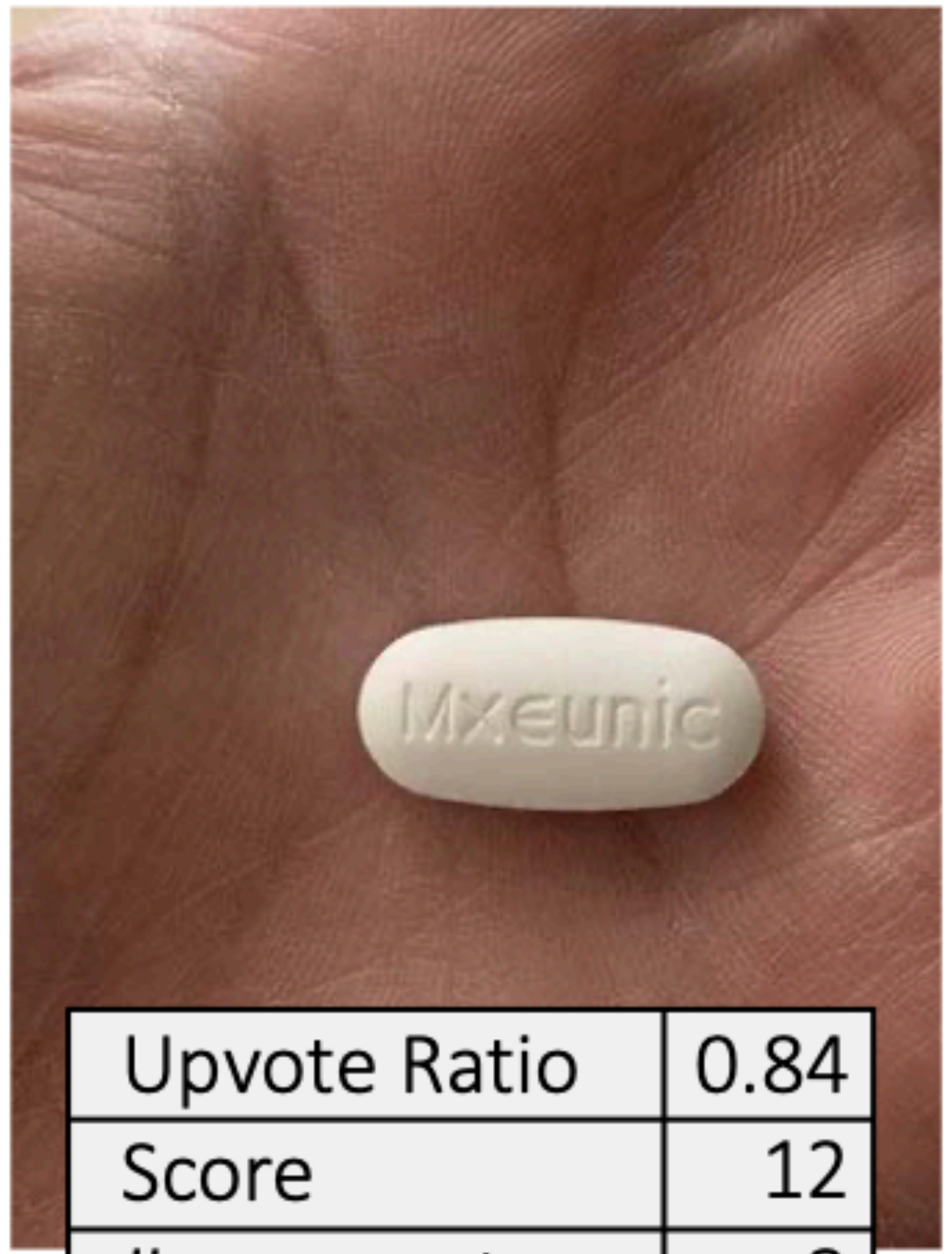
- Proposed a multi-stream network architecture that learns from four heterogeneous input modality, as well as metadata information.

Title:
The chickens hovering above the ground as well



Upvote Ratio	0.94
Score	4
#comments	60

Title:
My walgreens offbrand mucinex was engraved with the letters mucinex but in a different order



Upvote Ratio	0.84
Score	12
#comments	2

Introduction
Contribution

- Propose to fuse these four structurally different modalities at multiple levels to optimally account for the information contained in each modality.
- Investigate which modality is most important for the detection of information disorder and whether a combined multimodal analysis is beneficial in contrast to mono-modal processing.
- This approach leads to 2 conclusions:
    - All modalities can provide useful clues for the detection of fake news.
    - Proposed multilevel hierarchical information fusion allows to successfully capture information from all modalities.

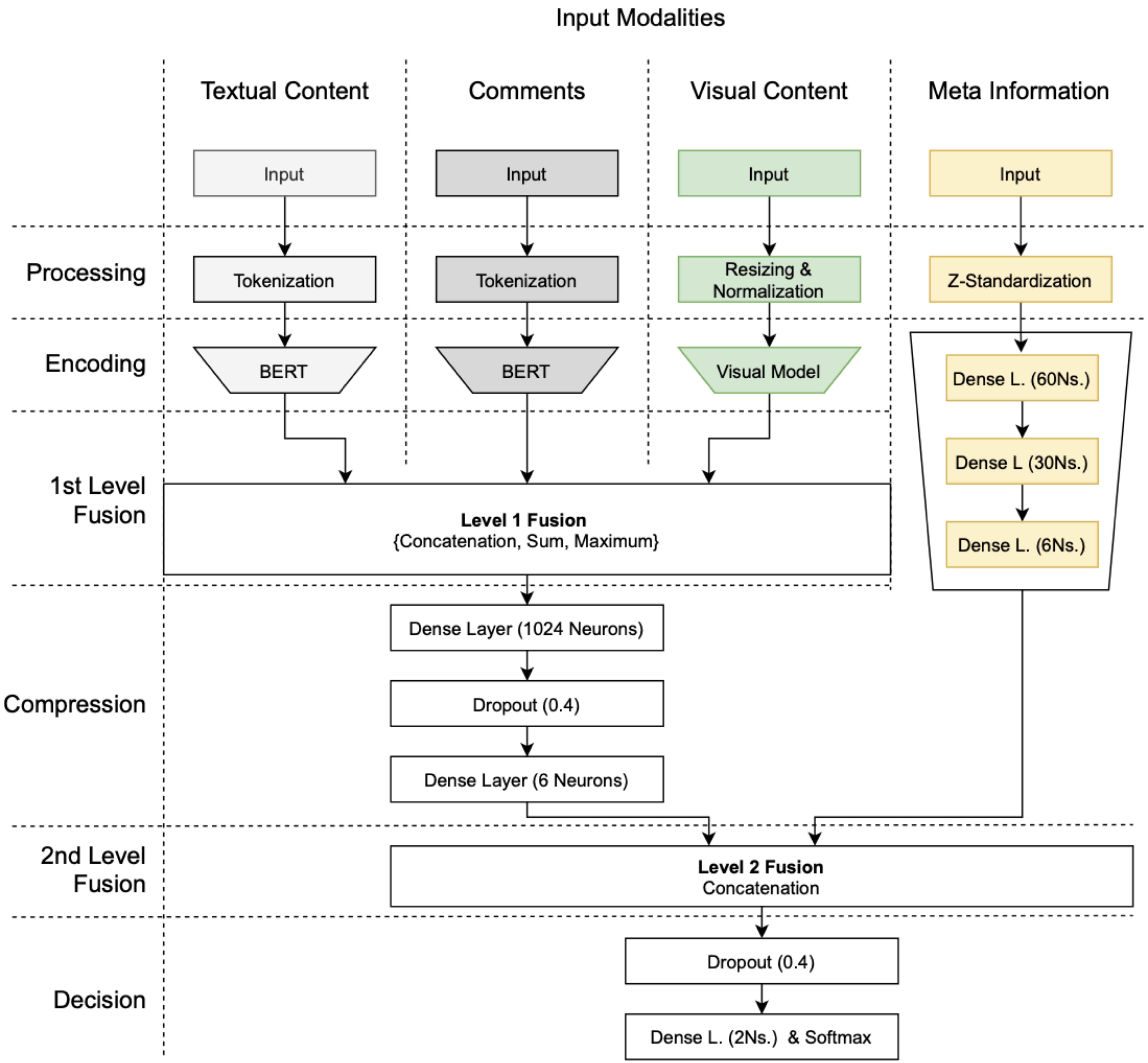
Related Work

of fake news detection

	Author	Textual Content	Visual Content	Metadata
RNN	Ma et al. (2018) [3]	x		
Stance	Mohtarami et al. (2018) [4]	x		
Image Trustworthiness	Lago et al. (2019) [5]		x	
CSI	Ruchansky et al. (2017) [10] CIKM	x		x
Social Rumor	Zubiaga et al. (2017) [9]	x		x
Dual	Dong et al. (2018) [8] WISE	x		x
EANN	Wang et al. (2018) [7] KDD	x	x	
SpotFAKE	Singhal et al. (2019) [6]	x	x	
r/fakeddit	Nakamura et al. (2020) [2]	x	x	
RNN	Jin et al. (2017) [12]	x	x	x
SAME	Cui et al. (2019) [11]	x	x	x
Videos.	Papadopoulou et al. (2019) [13]	x	x	x

Proposed Approach
Architectural Overview

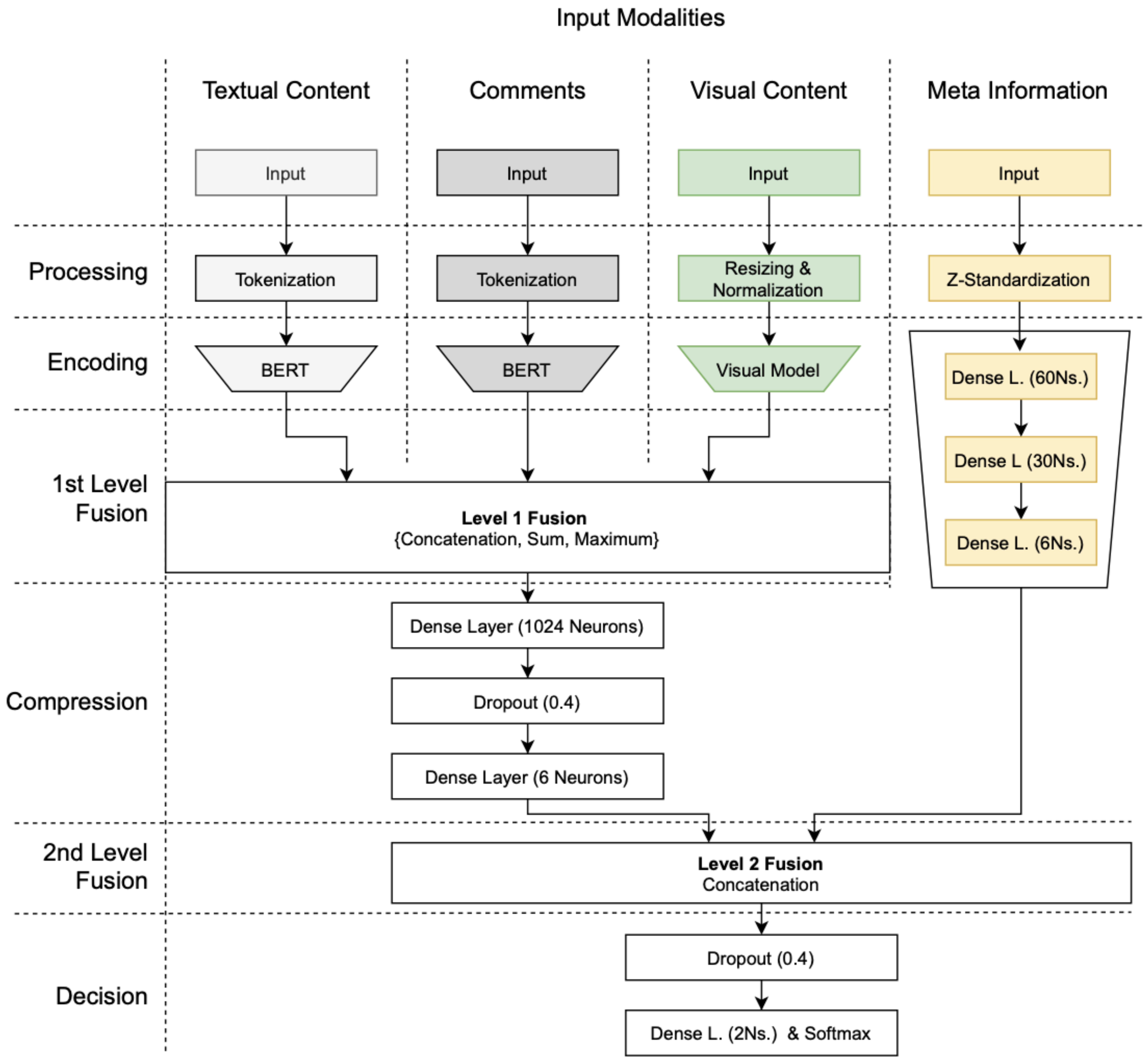
- Information disorder is a semantically complex concept that manifests itself in different modalities.
- Assume that the fusion of information from multiple modalities is important to solve this task.



Proposed Approach

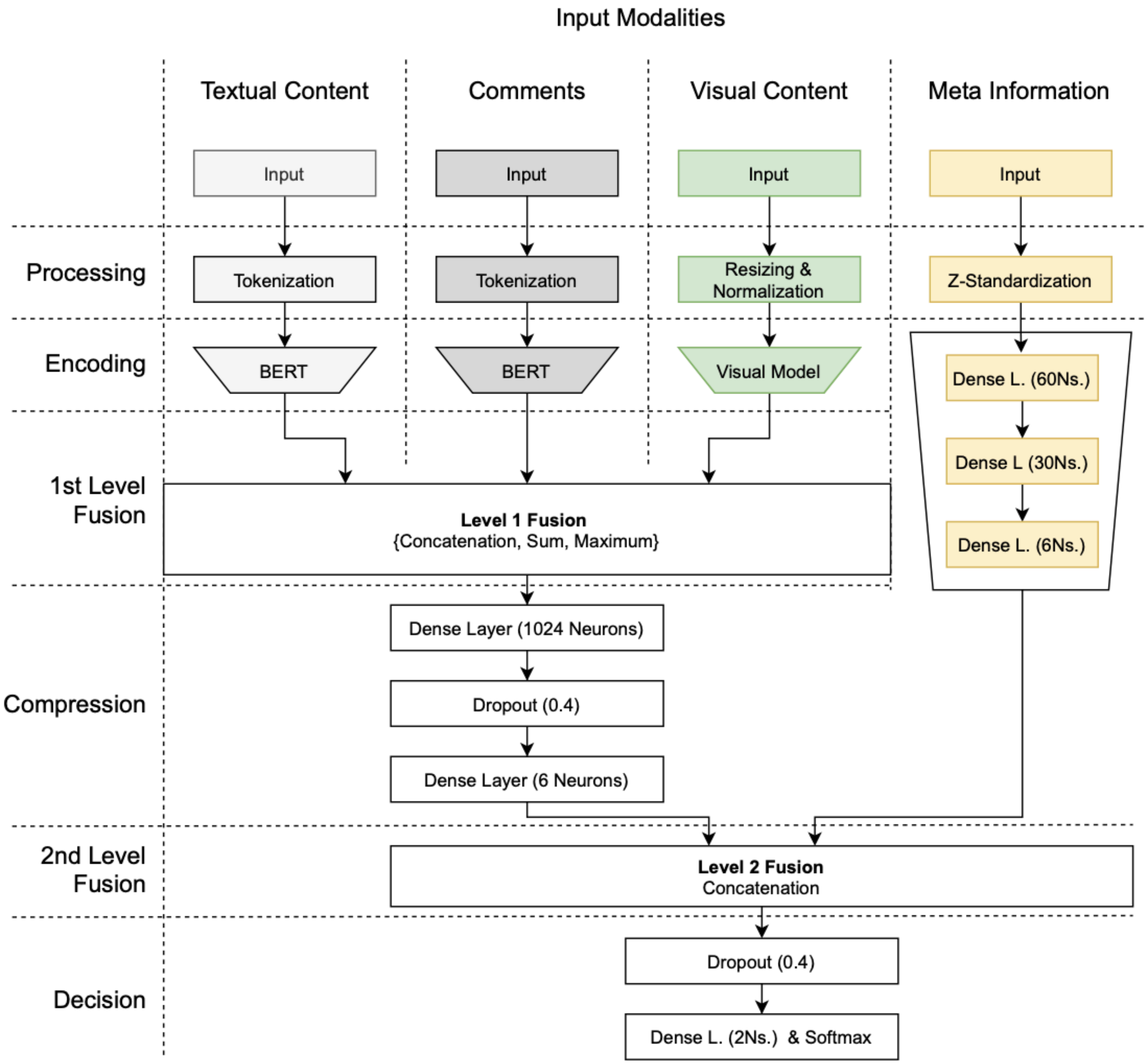
Architectural Overview

- Proposed an approach for information disorder detection based on four input modalities:
  - Primary textual content
  - Secondary information
  - Visual content of the posting
  - Available metadata info.



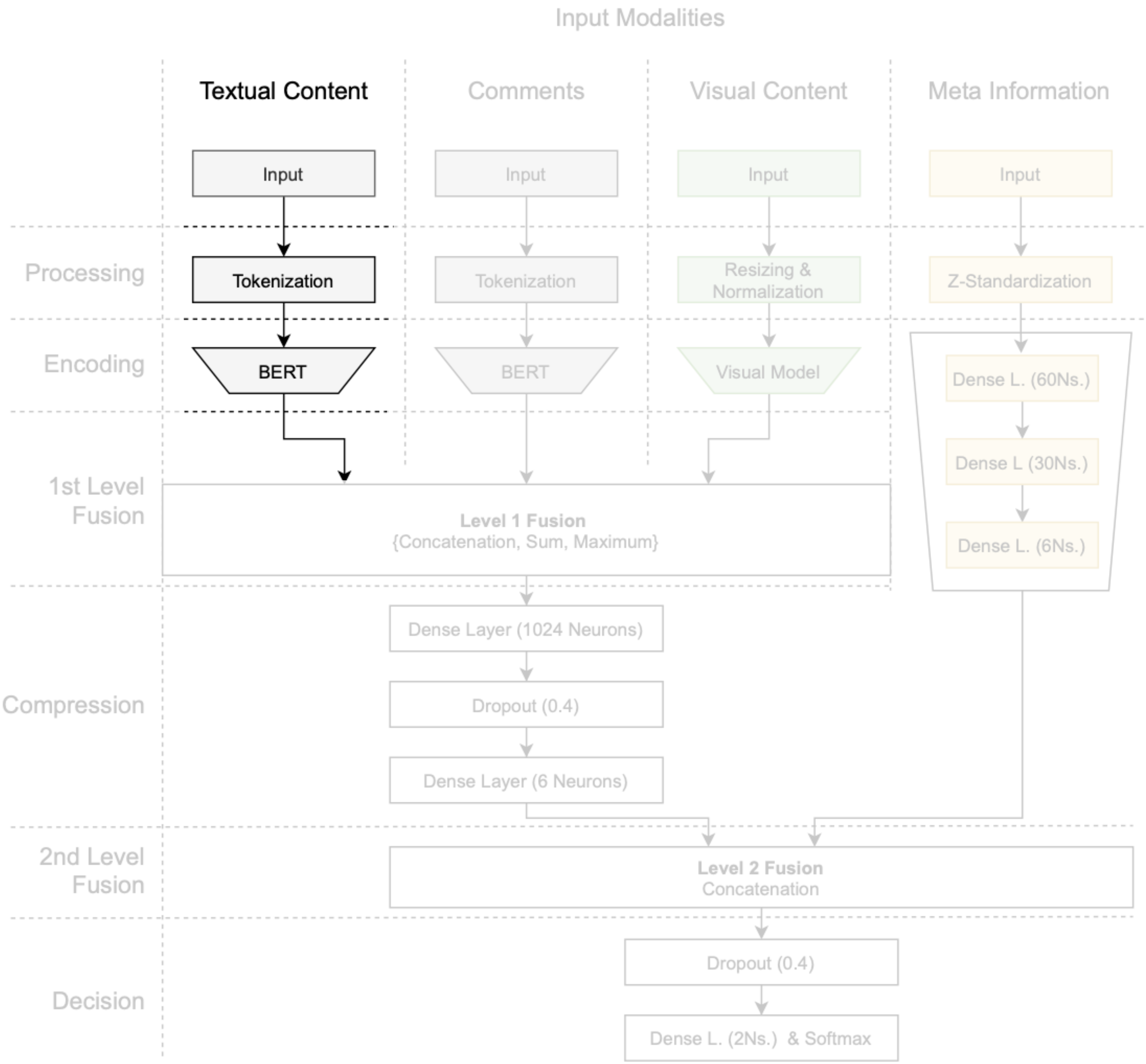
Proposed Approach
Architectural Overview

- A particular challenge is to fuse the information from these different types of input.
- Differ not only structurally but also in dimensionality.
    - Text vs. image
    - High-dimensional visual embedding vs. low-dimensional abstract data in case of metadata



Proposed Approach
Textual Content

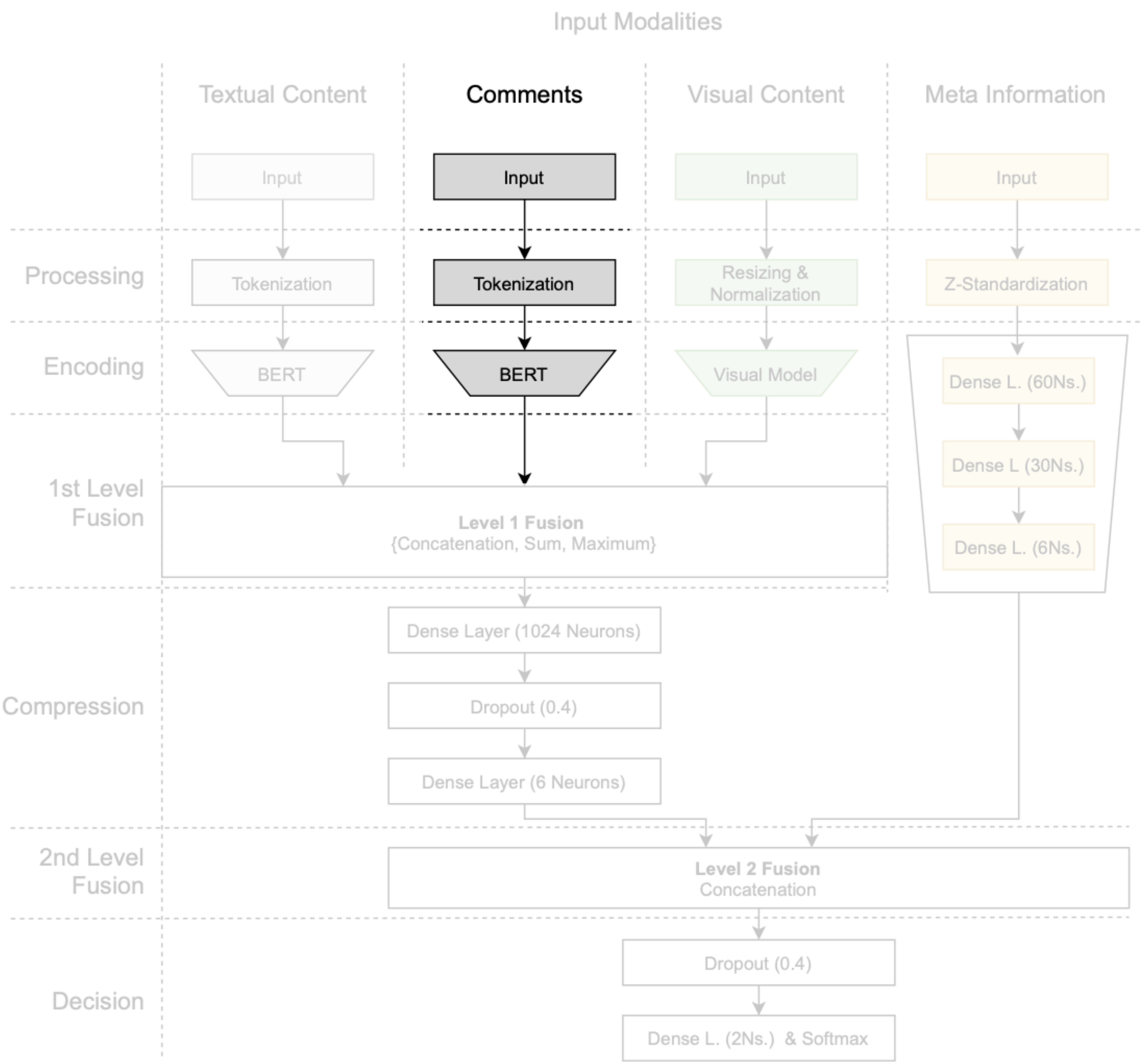
- The first stream takes the actual content of a social media posting as input.
- e.g. the title and, if available, its body.



Proposed Approach

Comments

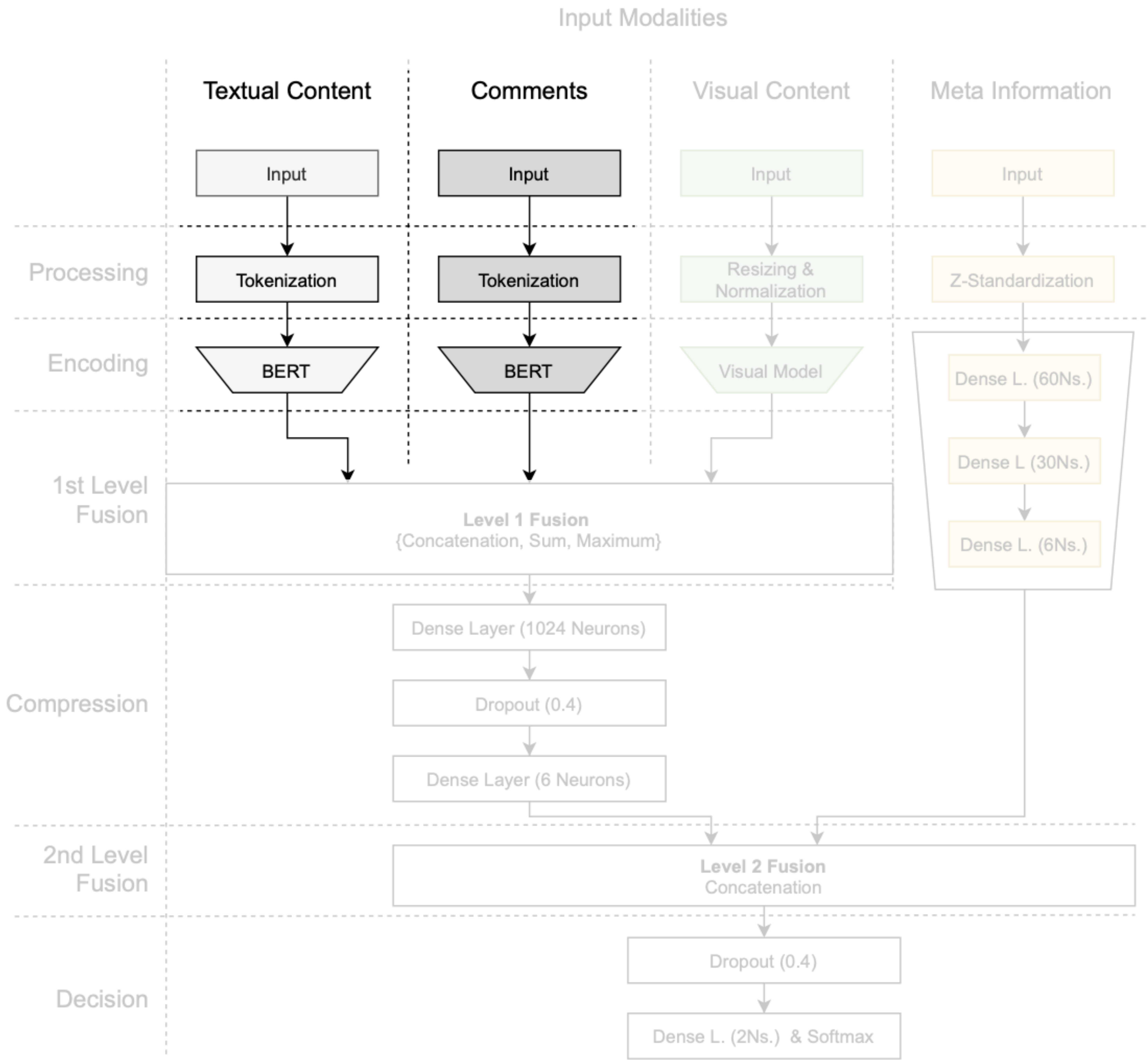
- Second stream processes textual information related to the posting.
- e.g. the comments available for the post.
- To keep the representation simple and comparable to the first stream, concatenate all available comments to obtain one consolidated input.



Proposed Approach

Process of textual modalities

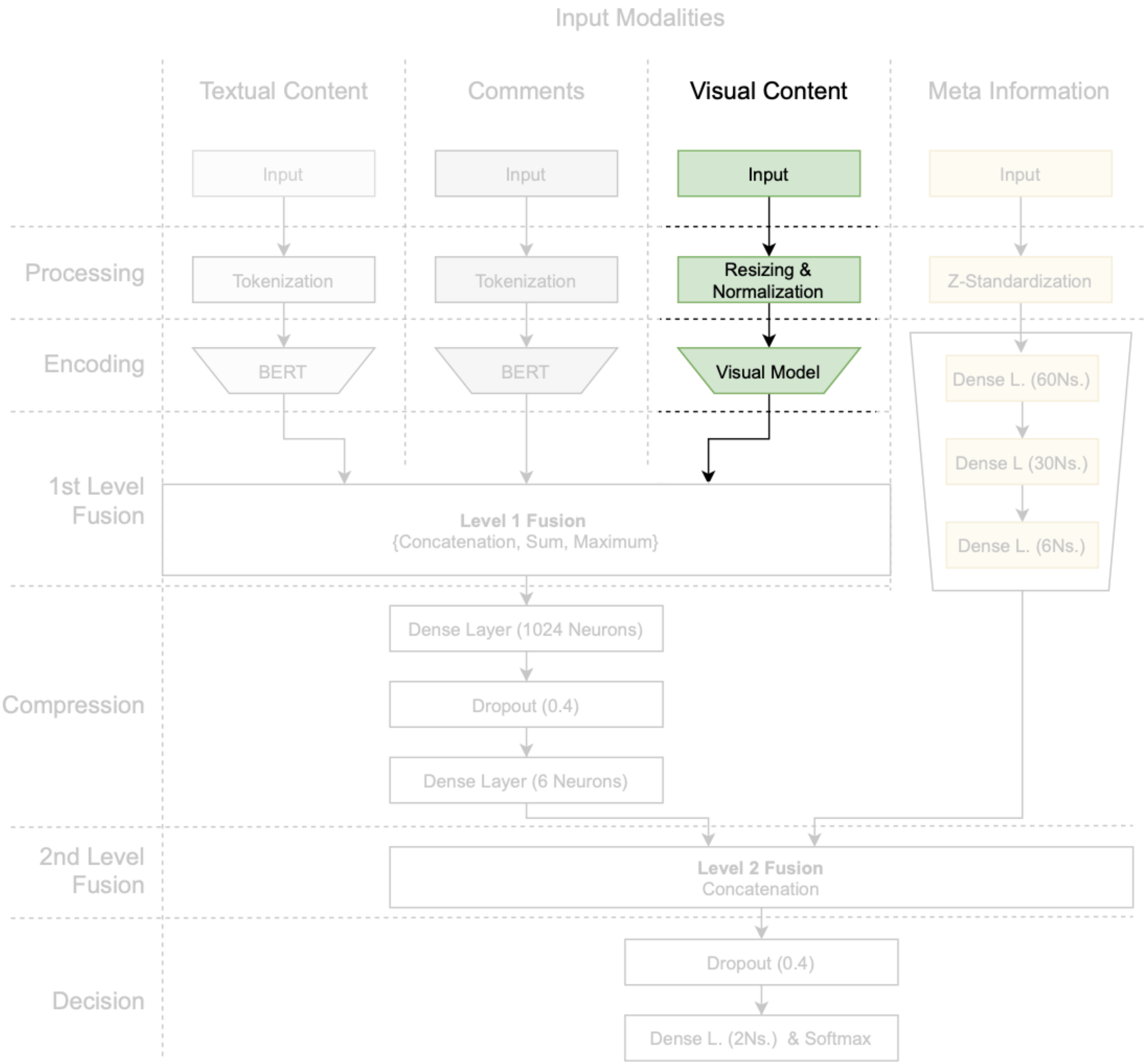
- Both textual modalities capture different perspectives on the actual content and are modeled in separate branches.
- Use a similar processing chain for both textual modalities.
- A BERT model is used to obtain separate text embeddings for the two inputs.



Proposed Approach

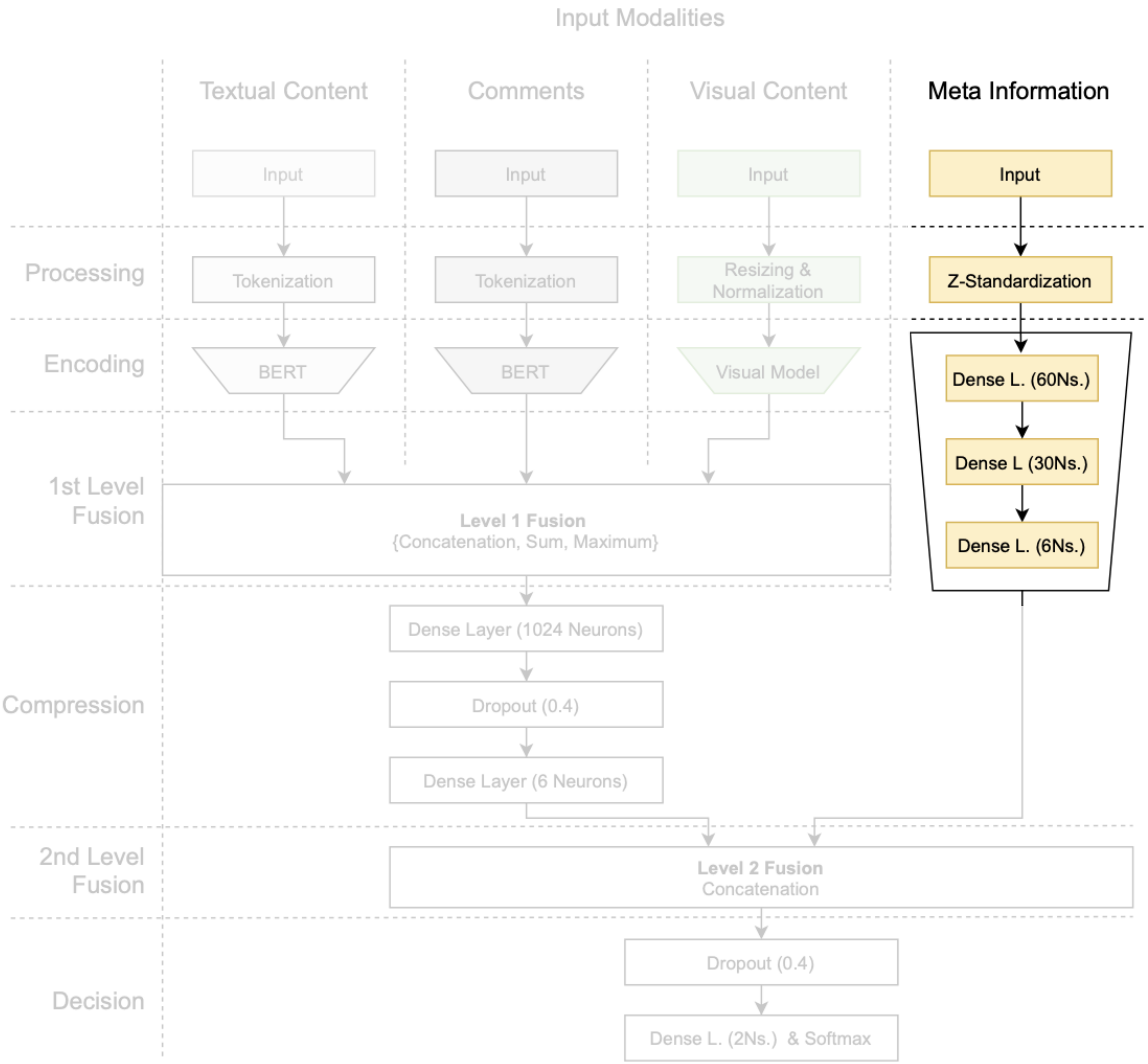
Visual Content

- First the images are standardized to zero-mean by calculating the mean over the entire training set (per channel) and subtracting it.
- After normalizing them to [0,1], the images are passed to a pretrained CNN to obtain a feature representation.
- e.g., ResNet, VGG



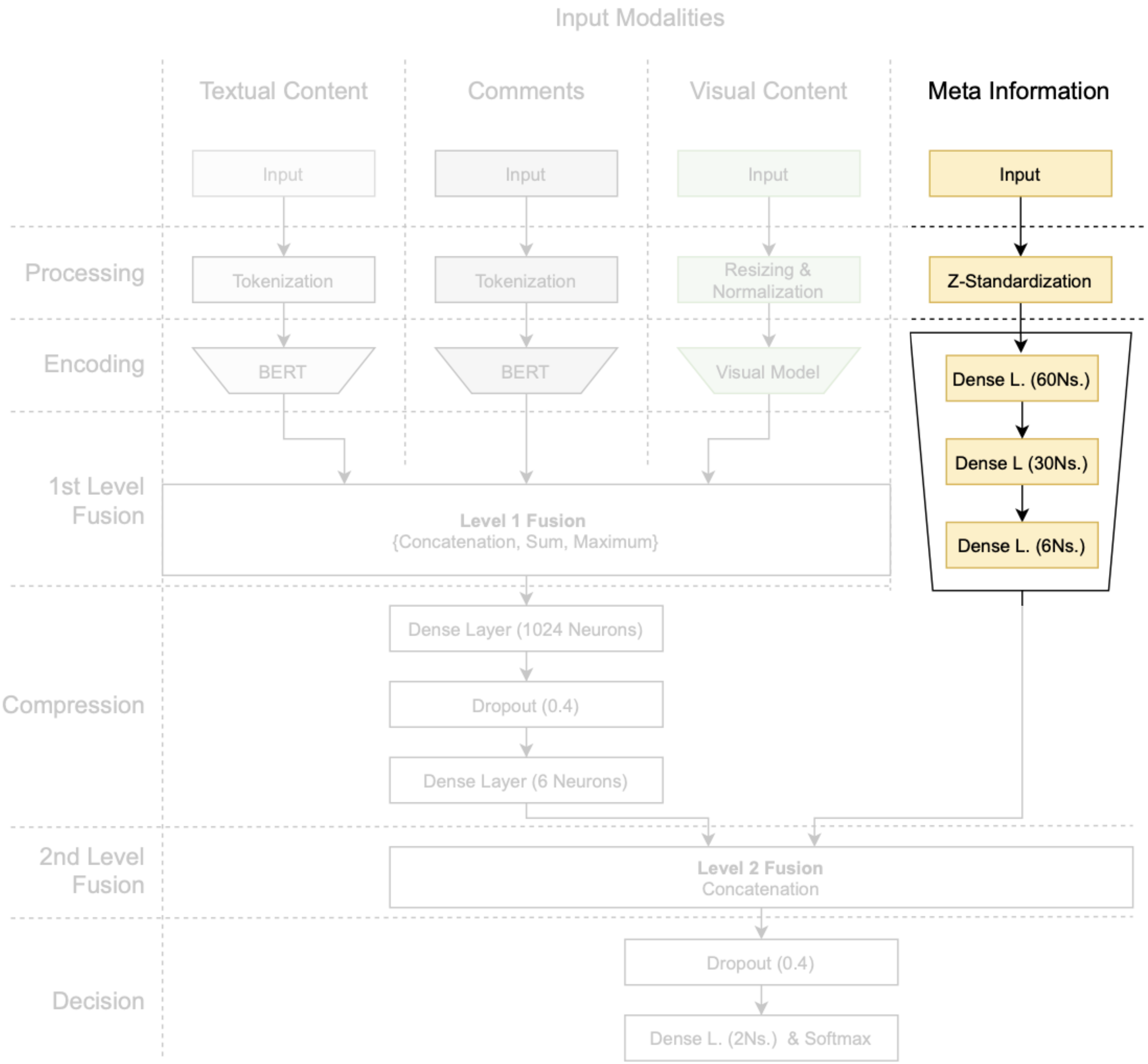
Proposed Approach
Meta Information

- Contain social media metrics or categorical data.
- e.g. the number of comments, the number of likes/dislikes, the number of upvotes or other ranking information.



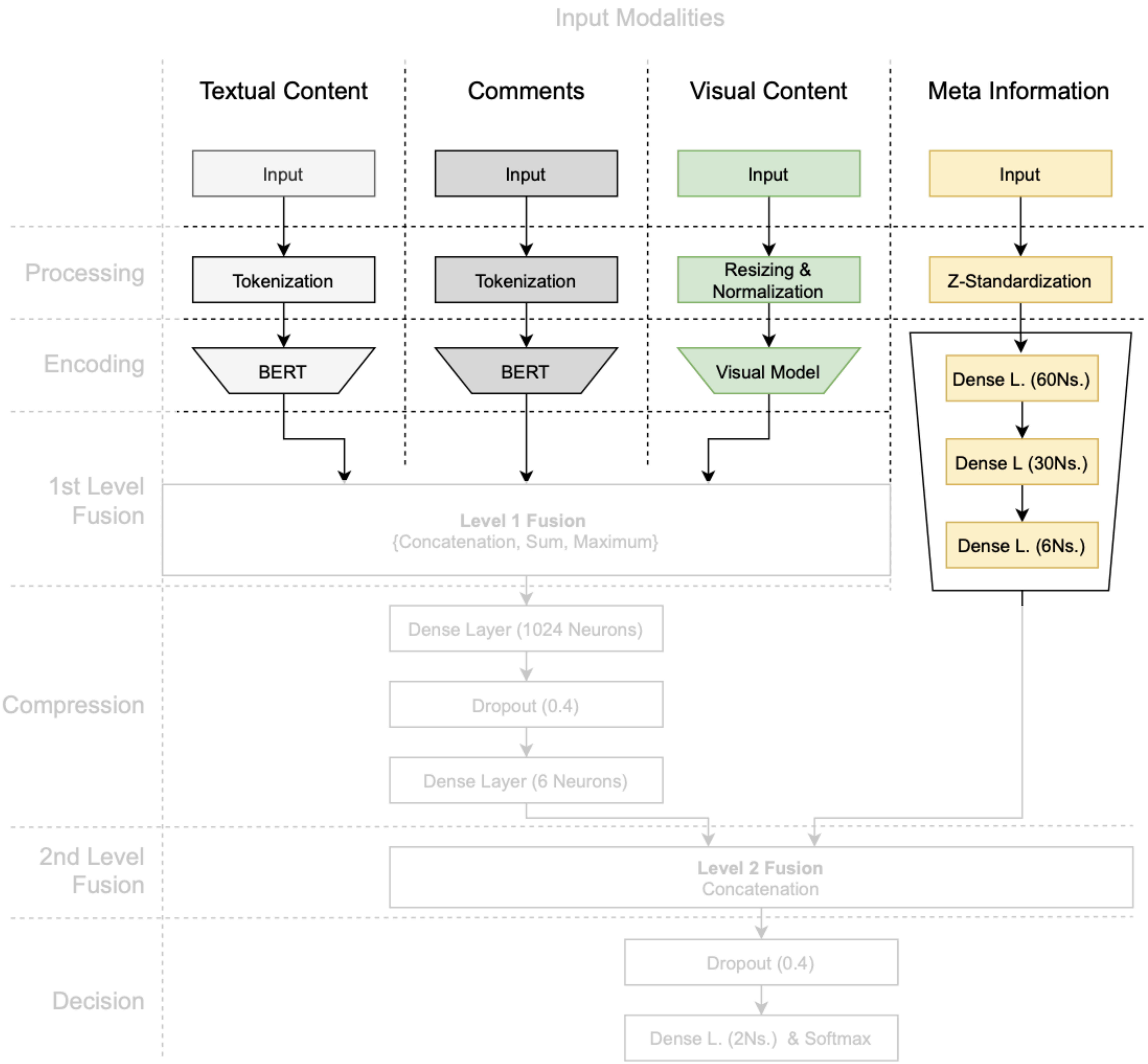
Proposed Approach
Meta Information

- First need to be normalized to a well-defined value range and then concatenated into a vector.
- Since no pre-defined encoder for such data exists, propose to train a lightweight multilayer perceptron (MLP) to represent the input data.
- Stack three dense layers and ReLU activation functions.



Proposed Approach
Fuse the information

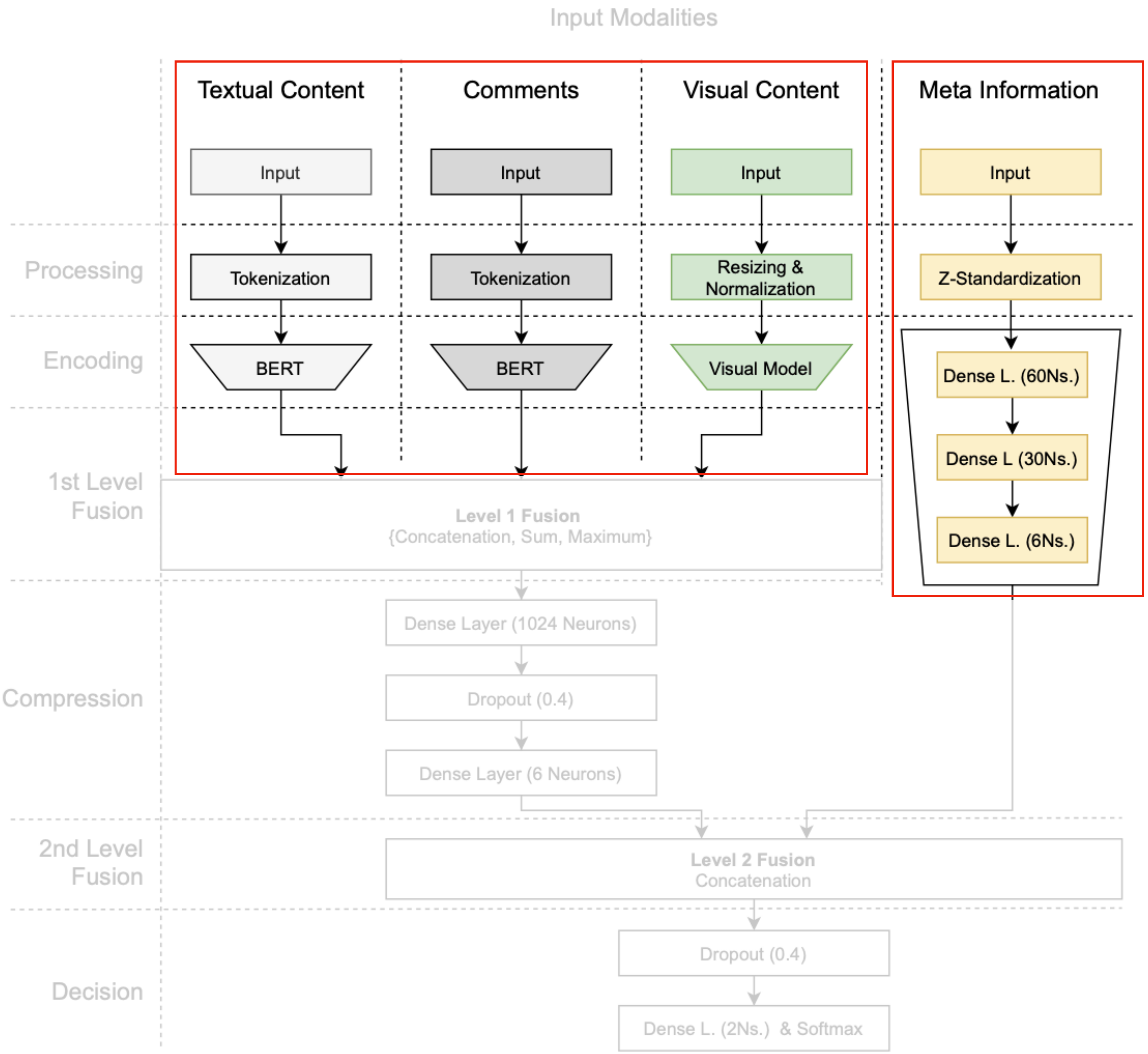
- Individual processing streams produce representations of different dimension.
- Thus propose a hierarchical scheme to fuse the information of the different modalities.



Proposed Approach

Fuse the information

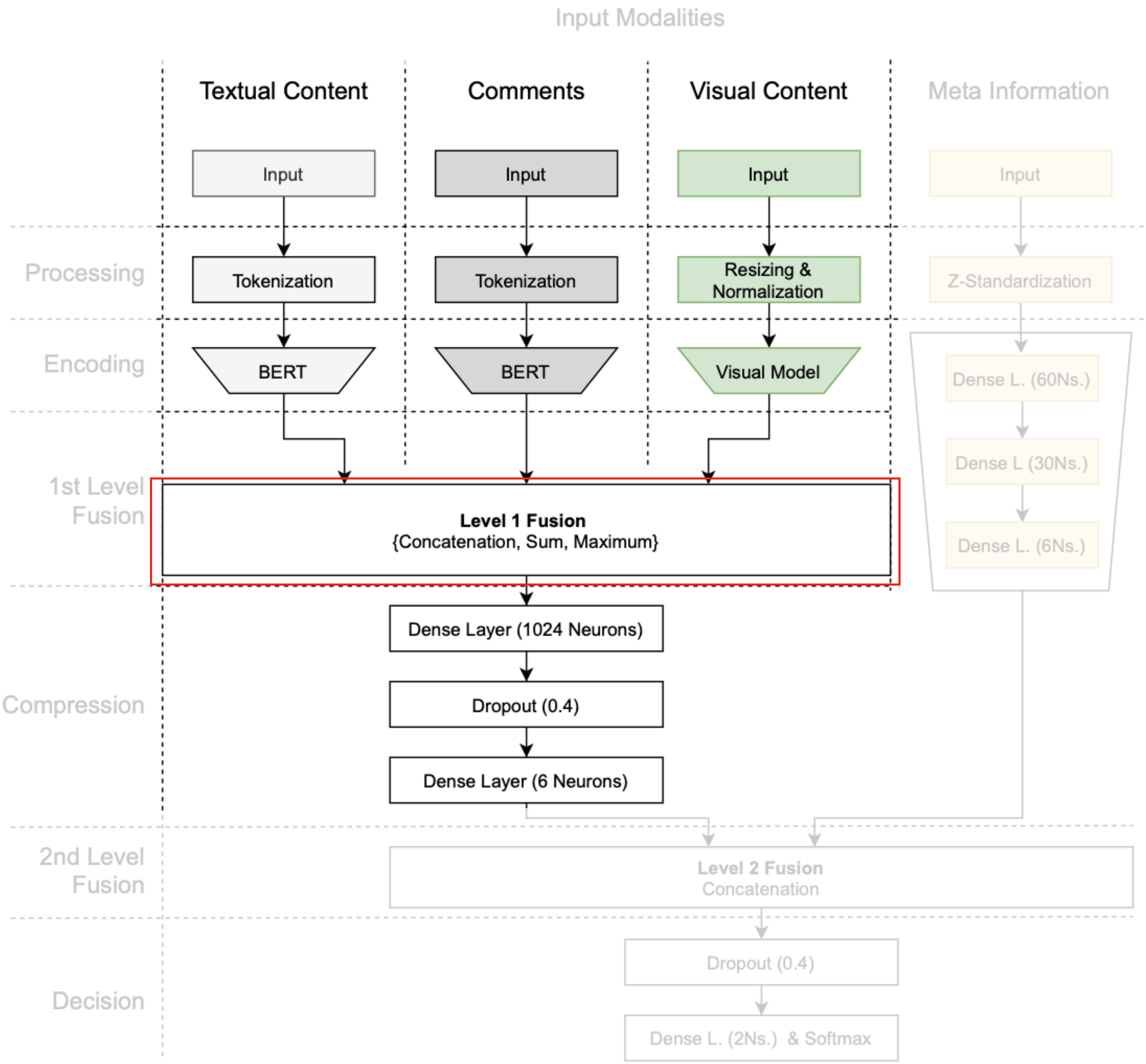
- This prevents that higher-dimensional representations dominate the other lower-dimensional representations like the one obtained from the metadata.



Proposed Approach

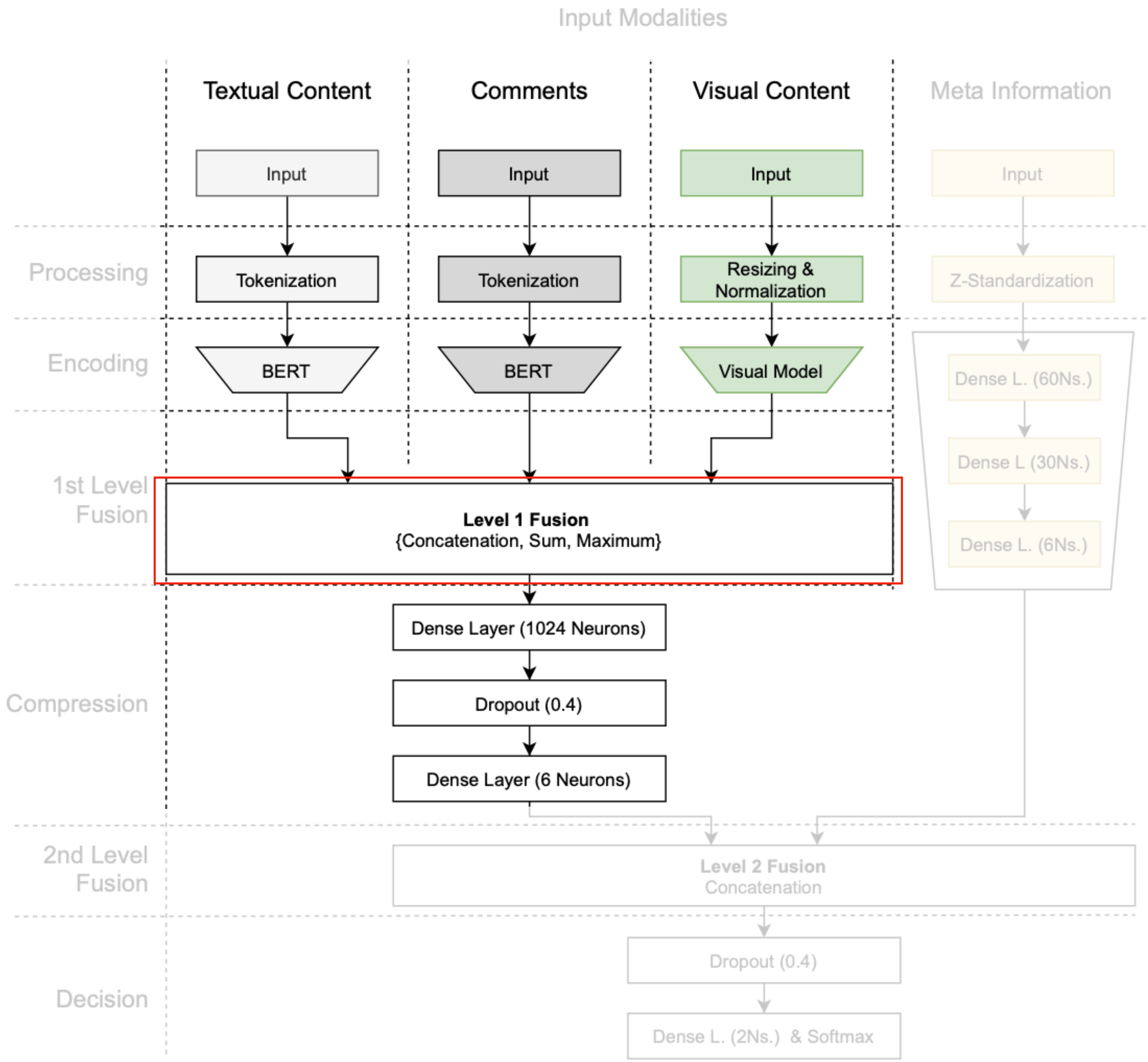
First level fusion

- Combines the textual and visual representations.
- These embedding vectors are designed to have all equal length (and thereby equal relevance in the fusion).



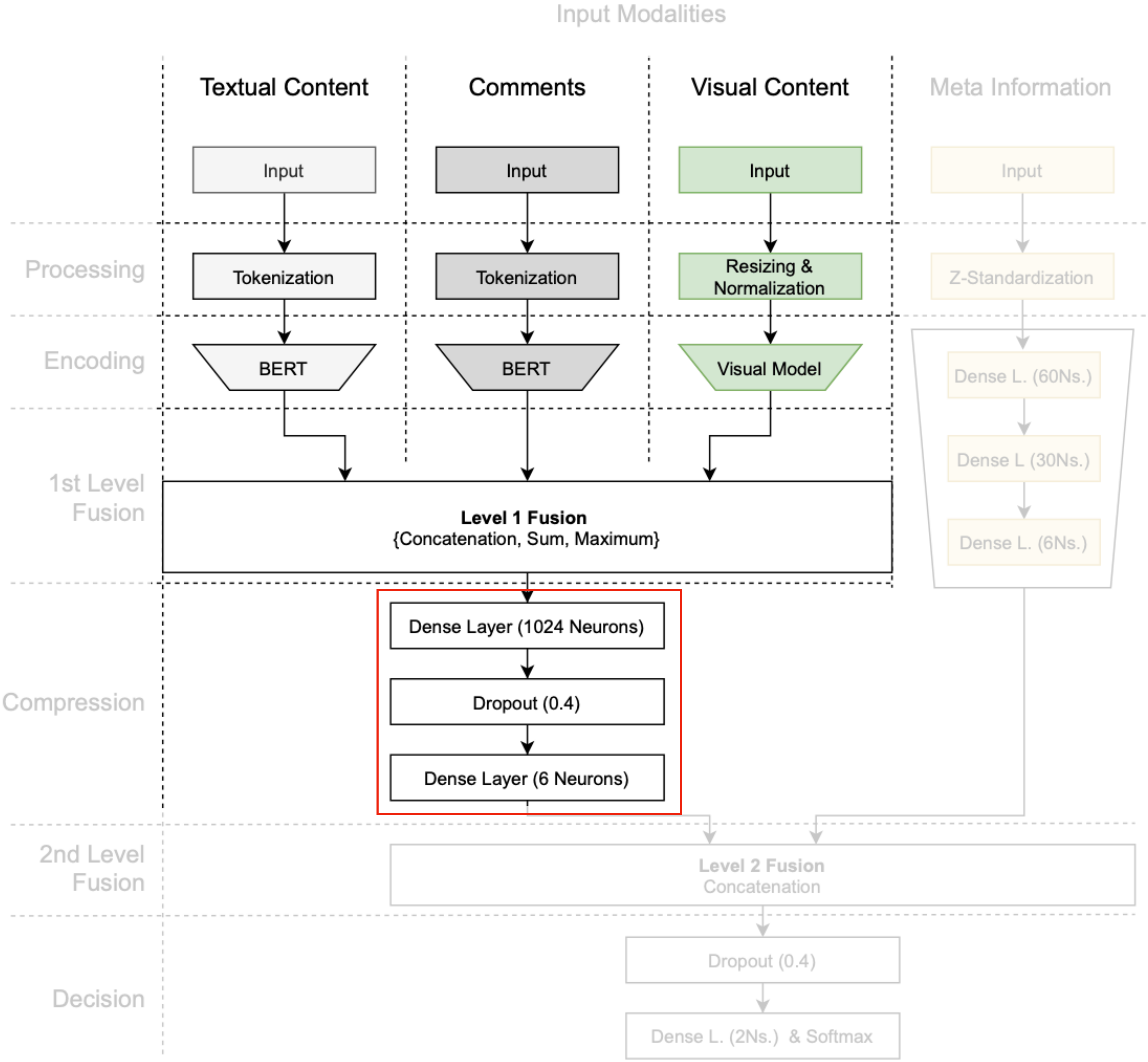
Proposed Approach
First level fusion

- This allows the use of different fusion strategies like concatenation, element-wise maximum of input vectors and element-wise average over all input vectors.
- Since it is not clear, which of these fusion operations is most beneficial, evaluate them systematically in experiments.



Proposed Approach
First level fusion

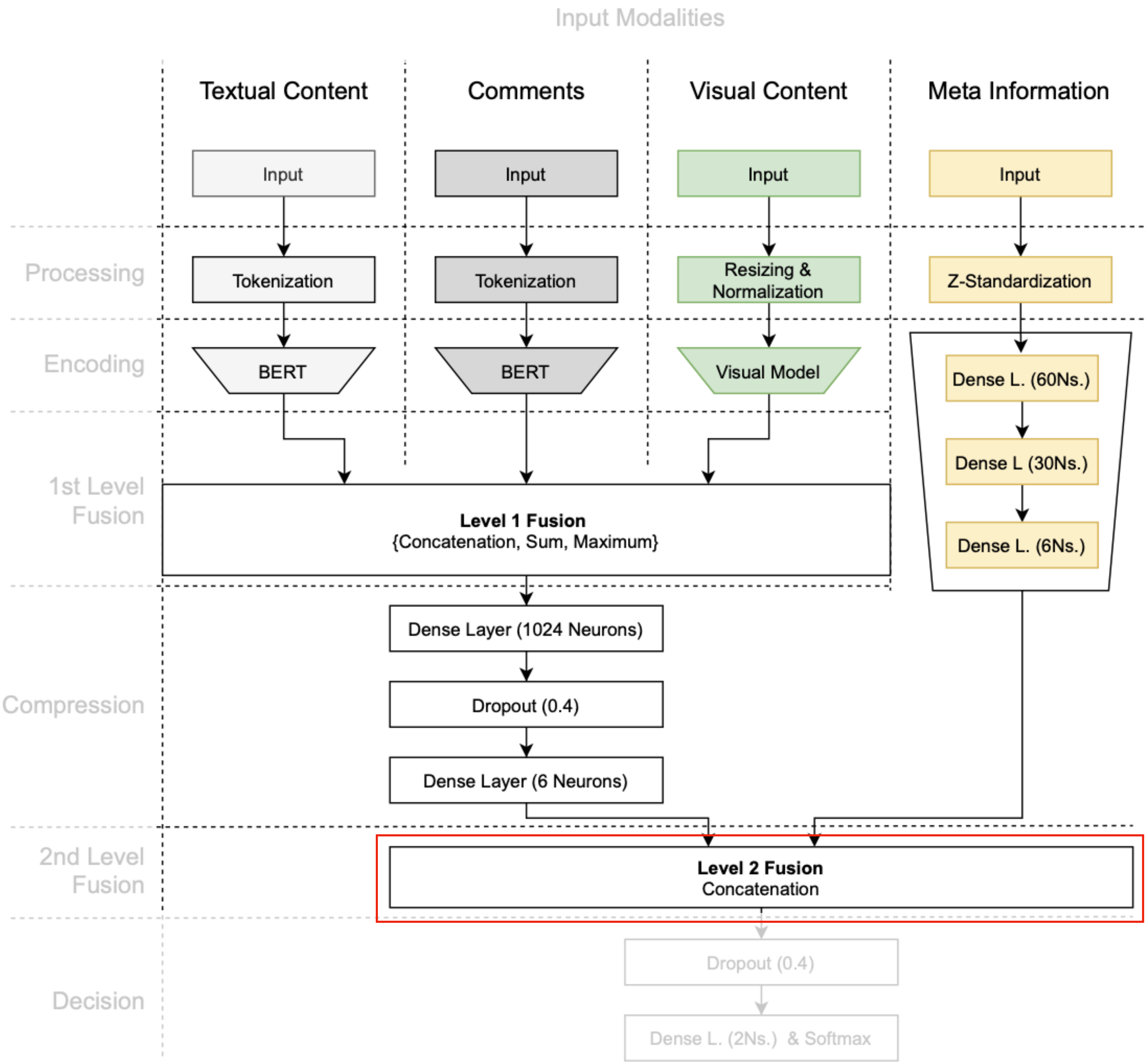
- The fused information is then further compressed by a stack of dense layers.
- So that it matches the dimensionality of the representation obtained by the fourth stream.



Proposed Approach

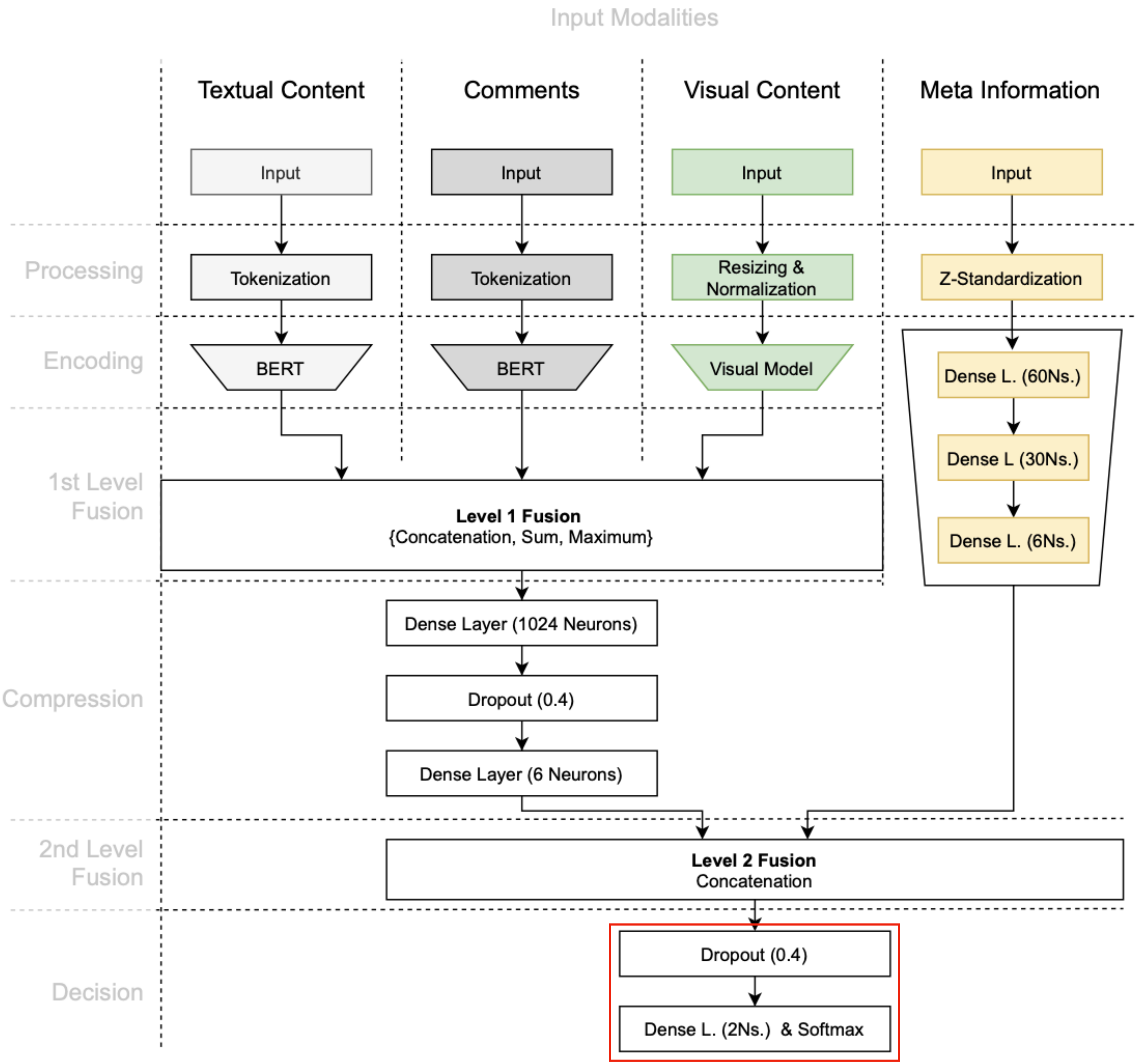
Second level fusion

- Two remaining representations are concatenated.
- Thereby, provide more influence to the metadata modality on the final detection (equal balance of content and metadata).



Proposed Approach
Final decision

- Final decision is made by a densely connected layer with two output neurons indicating fake vs. non-fake information.
- And followed by a softmax layer to obtain normalized probabilities.

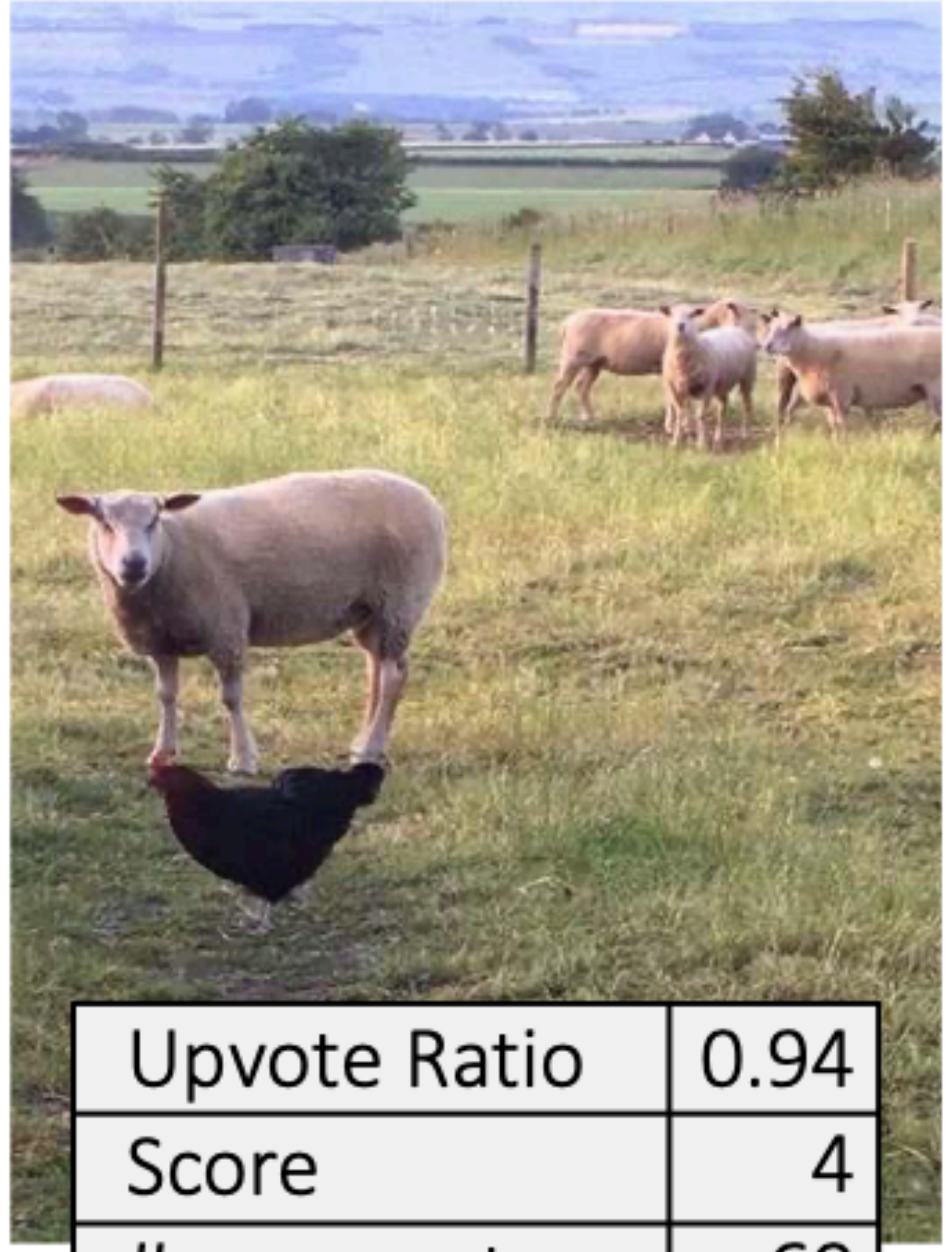


Experiments

Datasets

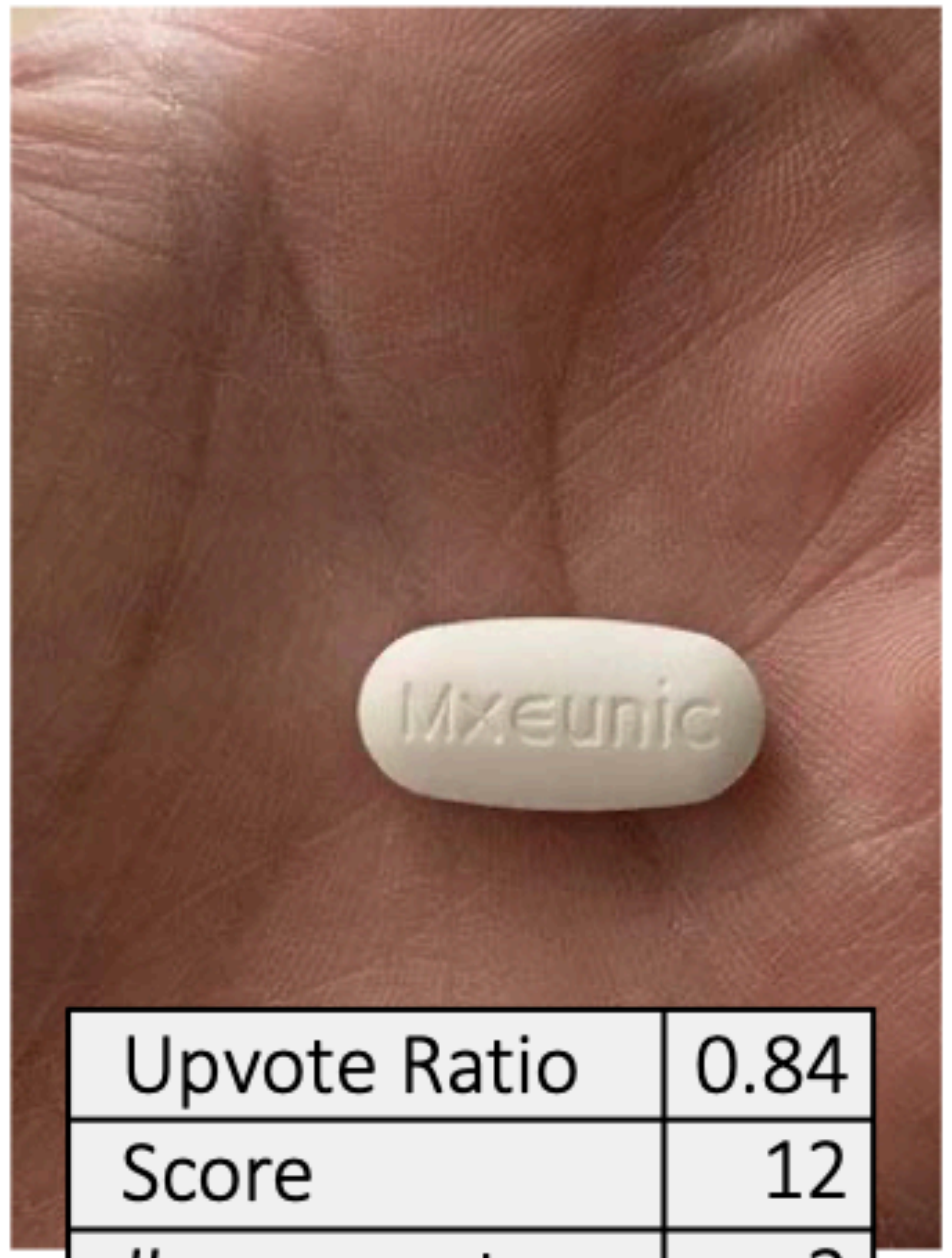
- Fakeddit dataset (LREC '20)
- The dataset contains Reddit postings with comments, with many of the postings contain text and images.
- Several metadata attributes like
  - up & downvotes of postings
  - the number of comments
  - up & downvote score for each comment
  - a score for the post itself

Title:
The chickens hovering above the ground as well



Upvote Ratio	0.94
Score	4
#comments	60

Title:
My walgreens offbrand mucinex was engraved with the letters mucinex but in a different order



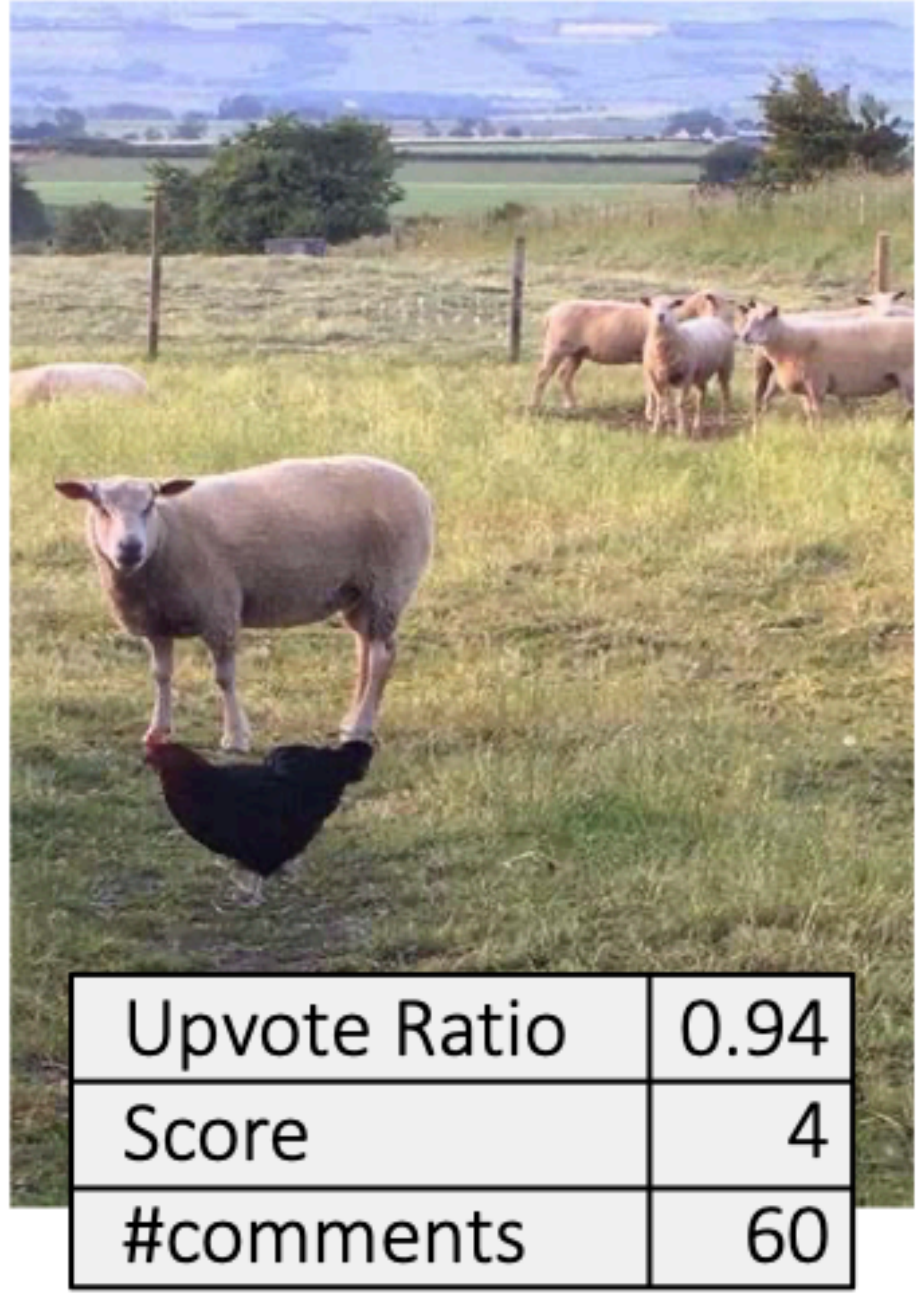
Upvote Ratio	0.84
Score	12
#comments	2

Experiments

Datasets

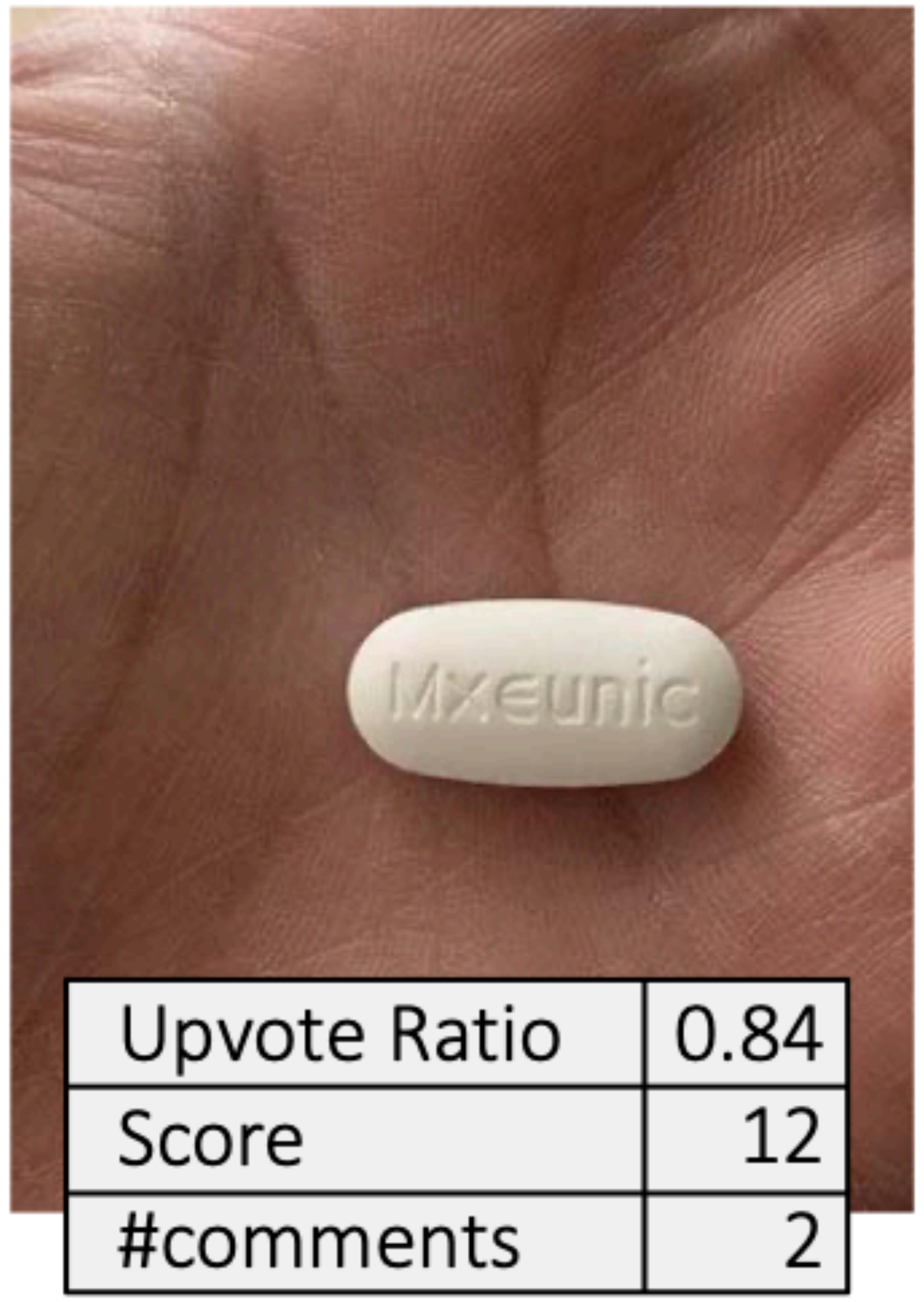
- Preprocess the data (similarly to r/Fakeddit) by removing samples where not all modalities are available (e.g. text-only postings).
- Results in
    - 560622 samples for training
    - 58972 samples for validation
    - 58954 holdout samples for testing.

Title:
The chickens hovering above the ground as well



Upvote Ratio	0.94
Score	4
#comments	60

Title:
My walgreens offbrand mucinex was engraved with the letters mucinex but in a different order



Upvote Ratio	0.84
Score	12
#comments	2

Experiments

Setup

- Textual data
  - Fed into the pre-trained BERT model
  - Sequence length of BERT is pre-allocated by shortening the input sequences to an average length (calculated over the training set) to reduce training time.
- Image data
  - Scaled and normalized fed into Inception-v3.
  - To assess the influence of different image resolutions, resize the images to 256x256px and 768x768px.

Experiments
Setup

- Metadata
  - Up & downvotes per post, its score and the count of comments.
  - To normalize the large value range of these attributes, z-standardize all metadata feature such as the count of comments and the score, except for the up & downvotes (already normalized between [0,1]).
  - The attributes are then provided to the three-layered MLP.

Experiments
Setup

- Training
  - Each modality can also been trained individually.
  - Achieved the best results by pre-training each modality (steam) separately, and then training only the fusion and classification layers on top.

Experiments
Baseline

- Use benchmark of r/Fakeddit dataset provide (LREC'20).
- Compare different fusion variants to estimate the best strategy for information fusion.
- Evaluate all possible combinations of modalities and further evaluate each modality in isolation to investigate the influence and expressiveness of each modality.

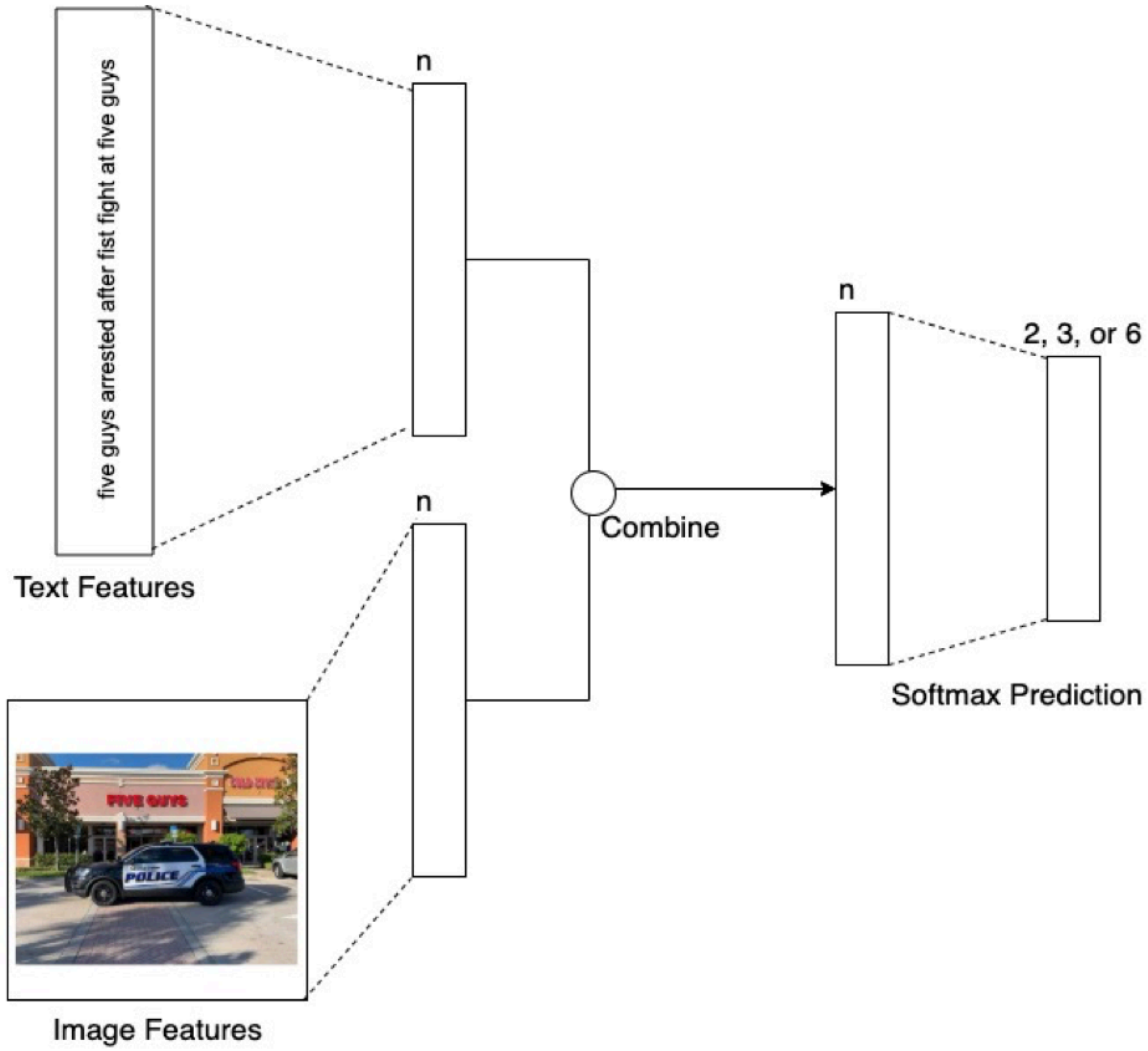


Figure 4: Multimodal model for integrating text and image data for 2, 3, and 6-way classification. n, the hidden layer size, is tuned for each model instance through hyperparameter optimization.

Experiments

Results

#	Approach	Textual Content	Textual Comments	Visual Content	Meta-data	Fusion Strategy	Val. Acc.	Test Acc.
1	Our approach	x	x	x	x	Sum	95.2%	95.5%
2	Our approach	x	x	x	x	Concat.	95.0%	95.2%
3	Our approach	x	x	x	x	Maximum	94.9%	95.1%
4	Our approach	x	x	x		Concat.	94.9%	95.0%
5	Our approach		x	x	x	Concat.	91.2%	91.3%
6	Our approach	x		x	x	Concat.	92.8%	92.8%
7	Our approach	x	x		x	Concat.	94.4%	94.5%
8	Our approach	x		x		Concat.	90.8%	91.0%
9	Our approach	x	x			Concat.	85.9%	85.7%
10	Our approach	x			x	Concat.	88.1%	88.2%
11	Our approach		x		x	Concat.	78.2%	78.2%
12	Our approach			x	x	Concat.	81.1%	81.6%
13	Our approach		x	x		Concat.	88.0%	88.1%
14	Our approach	x				-	88.1%	88.1%
15	Our approach		x			-	86.7%	86.5%
16	Our approach			x		-	81.0%	81.5%
17	Our approach				x	-	77.8%	77.3%
18	[2]	x				-	86.5%	86.4%
19	[2]			x		-	80.4%	80.7%
20	[2]	x		x		Maximum	89.3%	89.1%

- For individual modalities, observe that the most informative modality is the primary textual content, followed by secondary information (i.e. comments), the visual modality, and metadata.

Experiments
Results

#	Approach	Textual Content	Textual Comments	Visual Content	Meta-data	Fusion Strategy	Val. Acc.	Test Acc.
1	Our approach	x	x	x	x	Sum	95.2%	95.5%
2	Our approach	x	x	x	x	Concat.	95.0%	95.2%
3	Our approach	x	x	x	x	Maximum	94.9%	95.1%
4	Our approach	x	x	x		Concat.	94.9%	95.0%
5	Our approach		x	x	x	Concat.	91.2%	91.3%
6	Our approach	x		x	x	Concat.	92.8%	92.8%
7	Our approach	x	x		x	Concat.	94.4%	94.5%
8	Our approach	x		x		Concat.	90.8%	91.0%
9	Our approach	x	x			Concat.	85.9%	85.7%
10	Our approach	x			x	Concat.	88.1%	88.2%
11	Our approach		x		x	Concat.	78.2%	78.2%
12	Our approach			x	x	Concat.	81.1%	81.6%
13	Our approach		x	x		Concat.	88.0%	88.1%
14	Our approach	x				-	88.1%	88.1%
15	Our approach		x			-	86.7%	86.5%
16	Our approach			x		-	81.0%	81.5%
17	Our approach				x	-	77.8%	77.3%
18	[2]	x				-	86.5%	86.4%
19	[2]			x		-	80.4%	80.7%
20	[2]	x		x		Maximum	89.3%	89.1%

- The text-only and image-only (rows 14, 16) configuration outperform the respective configurations of (rows 18-19), therefore, represent new performance baselines.

Experiments
Results

#	Approach	Textual Content	Textual Comments	Visual Content	Meta-data	Fusion Strategy	Val. Acc.	Test Acc.
1	Our approach	x	x	x	x	Sum	95.2%	95.5%
2	Our approach	x	x	x	x	Concat.	95.0%	95.2%
3	Our approach	x	x	x	x	Maximum	94.9%	95.1%
4	Our approach	x	x	x		Concat.	94.9%	95.0%
5	Our approach		x	x	x	Concat.	91.2%	91.3%
6	Our approach	x		x	x	Concat.	92.8%	92.8%
7	Our approach	x	x		x	Concat.	94.4%	94.5%
8	Our approach	x		x		Concat.	90.8%	91.0%
9	Our approach	x	x			Concat.	85.9%	85.7%
10	Our approach	x			x	Concat.	88.1%	88.2%
11	Our approach		x		x	Concat.	78.2%	78.2%
12	Our approach			x	x	Concat.	81.1%	81.6%
13	Our approach		x	x		Concat.	88.0%	88.1%
14	Our approach	x				-	88.1%	88.1%
15	Our approach		x			-	86.7%	86.5%
16	Our approach			x		-	81.0%	81.5%
17	Our approach				x	-	77.8%	77.3%
18	[2]	x				-	86.5%	86.4%
19	[2]			x		-	80.4%	80.7%
20	[2]	x		x		Maximum	89.3%	89.1%

- By combining the two content modalities (text and images), baseline (row 20) yield a test accuracy of 89.1%.
- Proposed approach using the same modalities (row 8) yields 91%.
    - Note that it's the best result obtained by using just two modalities.

Experiments

Results

#	Approach	Textual Content	Textual Comments	Visual Content	Meta-data	Fusion Strategy	Val. Acc.	Test Acc.
1	Our approach	x	x	x	x	Sum	95.2%	95.5%
2	Our approach	x	x	x	x	Concat.	95.0%	95.2%
3	Our approach	x	x	x	x	Maximum	94.9%	95.1%
4	Our approach	x	x	x		Concat.	94.9%	95.0%
5	Our approach		x	x	x	Concat.	91.2%	91.3%
6	Our approach	x		x	x	Concat.	92.8%	92.8%
7	Our approach	x	x		x	Concat.	94.4%	94.5%
8	Our approach	x		x		Concat.	90.8%	91.0%
9	Our approach	x	x			Concat.	85.9%	85.7%
10	Our approach	x			x	Concat.	88.1%	88.2%
11	Our approach		x		x	Concat.	78.2%	78.2%
12	Our approach			x	x	Concat.	81.1%	81.6%
13	Our approach		x	x		Concat.	88.0%	88.1%
14	Our approach	x				-	88.1%	88.1%
15	Our approach		x			-	86.7%	86.5%
16	Our approach			x		-	81.0%	81.5%
17	Our approach				x	-	77.8%	77.3%
18	[2]	x				-	86.5%	86.4%
19	[2]			x		-	80.4%	80.7%
20	[2]	x		x		Maximum	89.3%	89.1%

- Adding metadata (row 6) yields 92.8%
- Adding comments (row 4) pushes performance to approx. 95%.
- The fusion of all 4 modalities (row 1-3) surpasses even the 95%.
- Observe that all three fusion strategies yield similarly good results.

Experiments
Results

#	Approach	Textual Content	Textual Comments	Visual Content	Meta-data	Fusion Strategy	Val. Acc.	Test Acc.
1	Our approach	x	x	x	x	Sum	95.2%	95.5%
2	Our approach	x	x	x	x	Concat.	95.0%	95.2%
3	Our approach	x	x	x	x	Maximum	94.9%	95.1%
4	Our approach	x	x	x		Concat.	94.9%	95.0%
5	Our approach		x	x	x	Concat.	91.2%	91.3%
6	Our approach	x		x	x	Concat.	92.8%	92.8%
7	Our approach	x	x		x	Concat.	94.4%	94.5%
8	Our approach	x		x		Concat.	90.8%	91.0%
9	Our approach	x	x			Concat.	85.9%	85.7%
10	Our approach	x			x	Concat.	88.1%	88.2%
11	Our approach		x		x	Concat.	78.2%	78.2%
12	Our approach			x	x	Concat.	81.1%	81.6%
13	Our approach		x	x		Concat.	88.0%	88.1%
14	Our approach	x				-	88.1%	88.1%
15	Our approach		x			-	86.7%	86.5%
16	Our approach			x		-	81.0%	81.5%
17	Our approach				x	-	77.8%	77.3%
18	[2]	x				-	86.5%	86.4%
19	[2]			x		-	80.4%	80.7%
20	[2]	x		x		Maximum	89.3%	89.1%

- The improvement over the baseline has two reasons:
  - Use two additional modalities that are useful for the task
  - Fine-tune all input streams (include BERT models), which alone yields around 2% performance gain.

Conclusion

- Proposed a multimodal architecture for the detection of information disorder, which incorporates not only the content of a social media postings but also metadata and secondary content related to the post.
- The additional modalities improve performance, indicate that they contribute useful information.
- Evaluation result shows that multimodal processing is superior to mono-modal processing.
- The authors plan to integrate a social network graph connecting postings, comments, and users as additional modality.

Comments

of Multimodal Detection of Information Disorder

- Using various types of modalities to detection fake news.
- Effective fusion strategy with high-low dimensional representation.
- Related work are present clearly and in recent years (17-20).
- Baseline method only compared with approach of dataset provide.
- May can improve by integrating with social network graph.