CSECU-DSG at SemEval-2021 Task 6: Orchestrating Multimodal Neural Architectures for Identifying Persuasion Techniques in Texts and Images

Tashin Hossain, Jannatun Naim, Fareen Tasneem,
Radiathun Tasnia, and Abu Nowshed Chy
Department of Computer Science and Engineering
University of Chittagong, Chattogram-4331, Bangladesh
{tashin.hossain.cu, jannatun.naim.cu, fareen.tasneem,
radia.tasnia.cu}@gmail.com and nowshed@cu.ac.bd

Abstract

Inscribing persuasion techniques in memes is the most impactful way to influence peoples' mindsets. People are more inclined to memes as they are more stimulating and convincing and hence memes are often exploited by tactfully engraving propaganda in its context with the intent of attaining specific agenda. This paper describes our participation in the three subtasks featured by SemEval 2021 task 6 on the detection of persuasion techniques in texts and images. We utilize a fusion of logistic regression, decision tree, and fine-tuned DistilBERT for tackling subtask 1. As for subtask 2, we propose a system that consolidates a span identification model and a multi-label classification model based on pre-trained BERT. We address the multi-modal multi-label classification of memes defined in subtask 3 by utilizing a ResNet50 based image model, DistilBERT based text model, and a multi-modal architecture based on multikernel CNN+LSTM and MLP model. The outcomes illustrated the competitive performance of our systems.

Keywords: persuasion techniques, transfer learning, multimodal neural architecture.

1 Introduction

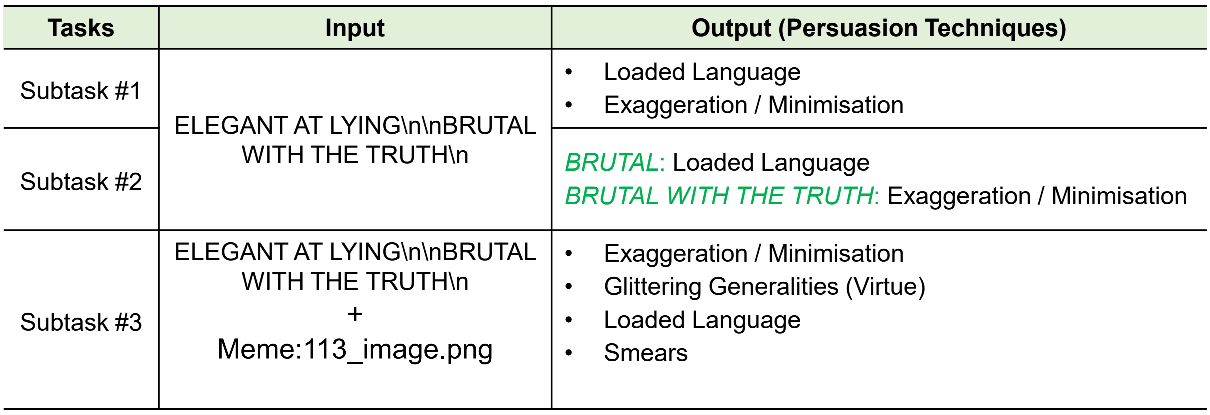
Persuasion techniques are quite recurrent in social media contents as it reaches a vast community. Proselytizing contents are adroitly implanted in posts and blogs which influence people's thoughts unconsciously. Nowadays such techniques are also being instilled in memes as people's attention is easily captured through illustration rather than narration. Manipulators often use this as a tool to promote their own deceitful agenda which can be political or anything else. Fake news is also spread through these disguised duplicitous contents which cause a lot of casualties. Therefore, it is an indispensable task to detect these techniques in multimodal contents to protect the users from deception.

The objective of SemEval 2021 task 6 (Dimitrov et al., 2021) is to detect the persuasion techniques in textual and multi-modal contents. This task includes three subtasks where the first two are based on textual contents only. More precisely, the first subtask requires us to detect which persuasion techniques among the given 20 techniques are inscribed in the textual content whereas the second subtask requires us to not only find which techniques are used but also to find the specific span of the text each technique corresponds to. The third subtask is a multi-modal multi-label classification problem where we need to identify which of the given 22 techniques are engraved both in the textual and visual content of the meme. An example from the provided dataset along with the desired output for three subtasks is depicted in Figure 1.

Numerous works have been done on the multilabel classification of text contents. (Chalkidis et al., 2019) depicted the pre-eminent impact of bidirectional GRU with label-wise attention in the legal domain. A consolidation of latent emotion memory (LEM) network and Bi-GRU was exploited for multilabel emotion classification (Fei et al., 2020). Besides, SemEval 2020 task 11 (Da San Martino et al., 2020) introduced two subtasks including span identification of propagandistic fragments in text content and technique classification of propagandistic fragments. The top-performing team (Morio et al., 2020) in the span identification subtask utilized several pre-trained language models for both subtasks. They also proposed an effective ensemble method with stacked generalization. The winning team (Jurkiewicz et al., 2020) of the technique classification subtask approached with an ensemble of RoBERTa based models and utilized RoBERTa-CRF archi-

The first four authors have equal contributions.

Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021), pages 1088–1095
Bangkok, Thailand (online), August 5–6, 2021. ©2021 Association for Computational Linguistics



Tasks	Input	Output (Persuasion Techniques)
Subtask #1	ELEGANT AT LYING\n\nBRUTAL WITH THE TRUTH\n	• Loaded Language • Exaggeration / Minimisation
Subtask #2		*BRUTAL*: Loaded Language *BRUTAL WITH THE TRUTH*: Exaggeration / Minimisation
Subtask #3	ELEGANT AT LYING\n\nBRUTAL WITH THE TRUTH\n + Meme:113_image.png	• Exaggeration / Minimisation • Glittering Generalities (Virtue) • Loaded Language • Smears

Figure 1: An illustration of the different subtasks.

tecture for the span identification subtask. (Wen et al., 2020) addressed a multi-label image classification problem by following human behavior pattern where labels and image features extracted by the ConvNet were projected to a common latent vector space to capture label correlation. (Song et al., 2018) used a deep multi-modal CNN method for multi-instance multi-label image classification.

In this paper, we present our approaches to address the challenges of identifying persuasion techniques in the textual and multimodal contents as defined in SemEval 2021 task 6. We exploit various kinds of approaches ranging from traditional statistical classifiers to the state-of-the-art deep learning architecture (e.g. multi-kernel CNN+LSTM, MLP, and ResNet50) and transformer models (e.g. BERT, DistilBERT, and FastBERT) in our proposed unified architecture.

We arrange the rest of the paper as follows: we explicate our proposed framework in Section 2. Section 3 enfolds the experimental details and comparative performance analysis. We analyze the performance of our models and also portray an analysis of erroneous detection in Section 3.4. Finally, we conclude this paper with some future prospects in Section 4.

2 Proposed Architecture

2.1 Subtask 1: Multi-label Persuasion Techniques Classification

In subtask 1, we need to design a method to identify the persuasive techniques used in textual content of a meme. The overview of our proposed system is depicted in Figure 2. In our proposed system, we combine three different models: 1) Logistic regression classifier, 2) Decision tree classifier, and 3) Fine-tuned DistilBERT model. We apply some preprocessing techniques including removing punctuations, numbers, special and single characters, multi-space, text lower-casing, word contradiction, and lemmatizing (Loper and Bird, 2002). Using our proposed models, we get different probability values for corresponding labels. Comparing our threshold score against the probability values, we find multi-label predictions from the individual models and employ the majority voting scheme to obtain our final multi-label predictions.

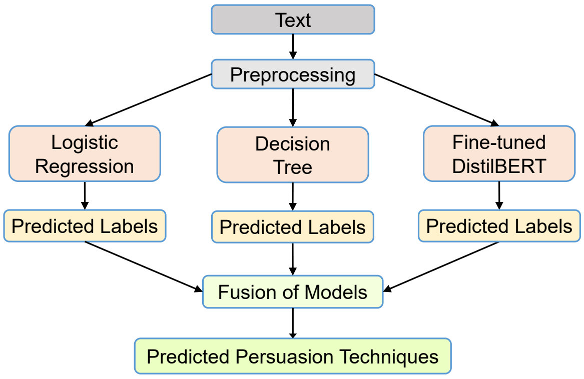


Figure 2: Proposed framework of Subtask 1.

2.1.1 Logistic Regression

Logistic regression (Cheng and Hüllermeier, 2009) is a machine learning model using probability concepts. It exploits some set of discrete values and the result is converted into a probability score by using a logistic sigmoid function. In our system, we employ a Tf-Idf vectorizer scheme for effective feature representation. We fix our threshold score to 0.05 for converting the probability score into a specific label category. If the probability score is greater than threshold values, it returns 1 as a true value for a specific label and vice versa.

2.1.2 Decision Tree

The decision tree (Safavian and Landgrebe, 1991) is a supervised learning classifier where values are divided ceaselessly following some specific parameters. We divide the decision tree into two subcomponents, one is decision nodes which split our values and another is leaves which are considered as final decided outcomes. For multi-label classification, we get different probabilities for all the class labels and set the threshold value to select labels following the same process as employed in the logistic regression.

2.1.3 Fine-tuned DistilBERT

DistilBERT (Sanh et al., 2019) is a transformer model that has 40% fewer parameters than BERT-base and works 60% faster. We fine-tuned DistilBERT model using the training dataset. For training purposes, we format the pre-processed data into two columns. One column contains the pre-processed text, and the other column carries labels. We convert the labels using scikit-learn(Pedregosa et al., 2011) MultiLabelBinarizer. We construct a neural network named DistilBERTClass involving the DistilBERT model along with the dropout and linear layer on top of it. The dimension of the linear layer is 20 which is the number of labels given in our subtask. We train the model a couple of times by feeding our dataset and we get the probability of each label. We use a random threshold to select the final labels.

2.1.4 Fusion of Models

We assemble our three individual models through a majority voting scheme. In majority voting, we count the occurrences of labels from three distinct models. We append the labels with the frequency of 2 or more to the final list of labels. Therefore, we obtain our final list of persuasive techniques for a given meme text.

2.2 Subtask 2: Span Identification of Persuasive Techniques

We propose a system that integrates a span identification model and a multi-label classification model for this subtask. We exploit an approach based on pre-trained BERT (bert-base-uncased). We employ SemEval 2020 Task 11's (Da San Martino et al., 2020) propaganda dataset as an external corpus. The overview of our proposed model is depicted in Figure 3.

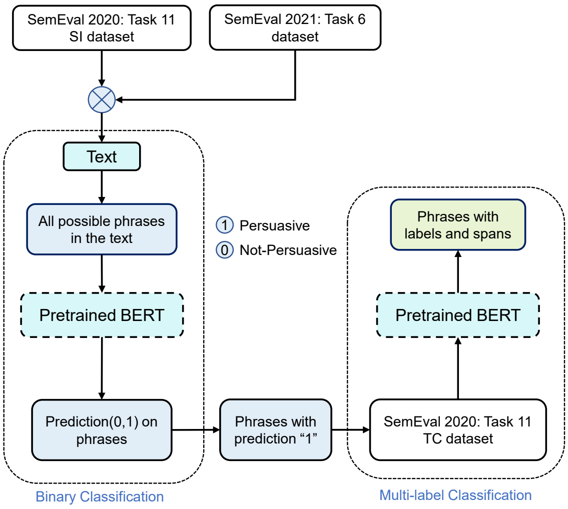


Figure 3: Proposed framework of Subtask 2.

2.2.1 Span Identification

We accumulate the sentences extracted from the articles of SemEval 2020 Task 11's SI dataset, SemEval 2021 Task 6's train, and development dataset. We derive all possible phrases from these sentences. Phrases with their indices included in span are labeled as 1 (Persuasive) while others are labeled as 0 (Not persuasive). This customized dataset is then sent to the pre-trained BERT model (Devlin et al., 2019) for training. We also extract all possible phrases from the test dataset. The pre-trained BERT model conducts binary classification on this test set. Here, the phrases are considered as sentences, so this process can be comprehended as a binary sentence classification task. After classifying the phrases derived from the test dataset, the indices of the phrases classified as 1 (Persuasive) are included in the spans and further processed for technique classification.

2.2.2 Technique Classification

The phrases of the test data that are predicted as persuasive in the previous segment are used as the test dataset of this segment. In this portion, we congregate SemEval 2020 Task 11's technique classification dataset, SemEval 2021 Task 6's train, and development dataset. In the case of the last two of them, we only include the text fragments, the indices of which are included in the provided spans instead of the whole text. We then send this contrived trainset to another pretrained BERT model with the same configuration as before and operate multi-label classification on the test set which generates predicted labels among the given 20 labels

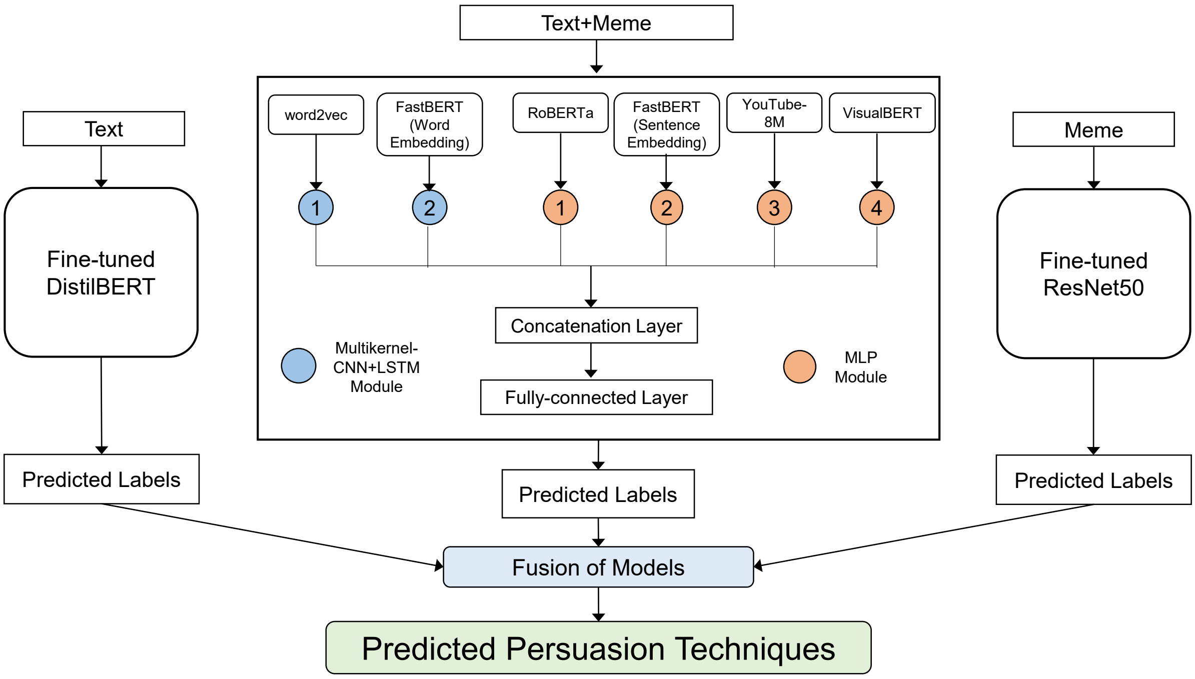


Figure 4: Proposed framework of Subtask 3.

per phrase. The phrases, their start index, end index, and their corresponding labels are then reintegrated as text fragments, start index, end index, and technique accordingly with their original text and converted into a suitable format for submission.

2.3 Subtask 3: Multi-modal Multi-label Classification

For this multi-modal subtask, we propose a majority voting based architecture as illustrated in Figure 4. We exploit a fine-tuned DistilBERT model, an ensemble of multi-kernel CNN with LSTM module and MLP module, and a fine-tuned ResNet50 model. These three models produce a list of persuasive techniques singularly and these lists are passed to the majority voting module to obtain the final list of persuasive techniques.

2.3.1 Fine-tuned DistilBERT

We use the same process of training described in Section 2.1.3. We accumulate the training and development dataset in a single corpus. Later, we use the 90% percent of the data for training and the rest of used as the validation set for finetuning.

2.3.2 Fine-tuned ResNet50

We perform fine-tuning on the residual neural network (He et al., 2016) having 50 layers. We convert our meme dataset as the format of the iMet Collection 2019 - FGVC6 dataset (Zhang et al., 2019). For training purposes, we include an additional label for the memes which have no labels assigned. We utilize the "ResNet50" pre-trained model, having "imagenet" as weights and 1000 classes. We interchange the Average pool layer with the AdaptiveAvgPool2d layer. We attach some batch normalization layers, dropout, and a linear layer. In the linear layer, the BatchNorm1d takes 2048 features as input. In the output layer, we return 23 output features where we add one additional label with the number of labels given in our problem. We train two layers such as layer4 and the last linear layer with the corresponding learning rate 1e-5 and 5e-3. We train the model numerous times and then get the model predictions. Finally, we set a random threshold to get the final predicted labels.

2.3.3 Ensemble of Multi-kernel CNN + LSTM and MLP Model

To address the challenge of the multimodal subtask, a combination of high-level features in a neural architecture is conventional. Our proposed model suggests a fusion of features extracted from multi-kernel CNN on top of the LSTM model and MLP (multi-layer perceptron) model. We exploit two kinds of word embeddings including word2vec (Mikolov et al., 2013) and fine-tuned FastBERT (Liu et al., 2020) models which are sent to the convolutional model of kernel size (2, 3) and subsequently to the LSTM model.

Besides, we also explore a multi-layer perceptron model for one-dimensional image features, sentence embeddings, and multi-modal features.

- **Image Features:** The image features are extracted from YouTube-8M (Abu-El-Haija et al., 2016) image feature extractor model(1024-dimension).
- **Sentence Embeddings:** These are extracted from the fine-tuned FastBERT (768-dimension) model and pre-trained RoBERTa (768-dimension) (Liu et al., 2019) model.
- **Multi-modal Features:** VisualBERT (Li et al., 2019) is exploited to blend image features along with text features. We implement the model proposed by (Li et al., 2020). We extract the image features utilizing Detectron2 (Wu et al., 2019) and the text features are encoded from a pre-trained BERT model. Both features are then merged inside VisualBERT. The dimension of the features is (164,768) and we flatten these features for our MLP module.

The output from two multi-kernel CNN+LSTM (MKCNN+LSTM) modules and four MLP modules are concatenated and further transmitted to the fully connected layer.

2.3.4 Fusion of Models

The list of predicted labels from the above three models are subsequently passed to a majority voting module. The primary idea behind majority voting is based on the frequency of the labels. If a label exists in the majority of the models, it is appended in the final list of labels.

3 Experiments and Evaluations

3.1 Dataset Description

In SemEval-2021 task 6 (Dimitrov et al., 2021), overall 950 data has been provided for subtask 1, 2, and 3. In the case of subtask 1 and 2, training, development, and test set contain 687, 63, and 200 textual data respectively. For subtask 3, the same amount of textual and meme data has been accommodated since it is a multi-modal subtask. Dataset for subtask 1 and 3 is annotated with 20 and 22 persuasive techniques correspondingly while subtask 2 dataset provides spans of 20 techniques used all together in the text.

3.2 Experimental Setup

In this section, we illustrate our submitted systems in SemEval-2021 Task 6. In case of subtask 1, we use three differents models i.e. logistic regression, decision tree classifier, and fine-tuned DistilBERT model. The system configuration of these three individual models are given in Table 1.

System	Settings
Logistic Regression	1. *max_iter*: 2000 2. *C*: 20 3. *penalty*: l2 4. *tol*: 0.001
Decision Tree	1. *min_samples_split*: 2 2. *min_samples_leaf*: 1 3. *criterion*: gini 4. *splitter*: best
Fine-tuned DistilBERT	1. *Tokenizer*: distilbert-base-uncased 2. *Dropout*: 0.2 3. *Learning rate*: 2e-5 4. *Batch size*: 16 5. *num_workers*: 4 6. *Maximum length*: 60 7. *Epochs*: 10

Table 1: System settings for Subtask 1.

We used the same system configuration of pre-trained BERT model for two segments i.e. span identification and multi-label technique classification in the subtask 2. The system settings are depicted in Table 2.

System	Settings
Pre-trained BERT	1. *max_seq_length*: 128 2. *Epochs*: 1 3. *train_batch_size*: 8 4. *eval_batch_size*: 8 5. *Weight decay*: 0.5 6. *Learning rate*: 4e-5 7. *adam_epsilon*: 1e-8 8. *warmup_ratio*: 0.06 9. *max_grad_norm*: 1.0 10. *gradient_accumulation_steps*: 1 11. *logging_steps*: 50 12. *save_steps*: 2000

Table 2: System settings for Subtask 2.

For subtask 3, we used three types of models. One is a fine-tuned DistilBERT model which is trained using the text written in a meme. The other model is a fine-tuned ResNet50 model, and the last one is multi-kernel CNN+LSTM and MLP model. These three models trained with the given dataset using different parameter settings. The system settings for each model are represented in Table 3. As meme is a combination of text and image, therefore we consider the majority voting based predictions as the final predictions for subtask 3.

System	Settings
MultiKernel LSTM with MLP	1. *nb_filters*: 200 2. *nb_rnnoutdim*: 600 3. *rnn_dropout*: 0.5 4. *optimizer*: adam 5. *Threshold*: 0.3 6. *Epochs*: 600
Fine-tuned DistilBERT	1. *Tokenizer*: distilbert-base-uncased 2. *Dropout*: 0.25 3. *Learning rate*: 1e-4 4. *Batch size*: 16 5. *maximum length*: 60 6. *Epochs*: 30
Fine-tuned ResNet50	1. *Image Size*: (224,224,3) 2. *Train Batch Size*: 32 3. *Test Batch Size*: 16 4. *Optimizer*: Adam 5. *Optimizer Learning rate*: 2e-4 6. *Epochs*: 900

Table 3: System settings for Subtask 3.

3.3 Results and Analysis

We now compare our proposed CSECUDSG system's performance with other participants systems in three subtasks as shown in Table 4, Table 5, and Table 6, respectively. In all the subtasks, the baseline system is set to random. The organizers used the F1-Micro as the primary evaluation measure for all the subtasks.

The overall scores of the three subtasks portray that our system acquired competitive performance. However, in all the subtasks, our system has some shortcomings with respect to the top performing teams. MinD, Volta, and Alpha are the top-performing teams in corresponding subtasks. We further analyze the performance of our systems in the subsequent section.

Team_Name	F1-Macro	F1-Micro
MinD	0.28993	0.59331
Alpha	0.26218	0.57187
Volta	0.26621	0.56958
CSECUDSG	**0.18454**	**0.48894**
NLPIITR	0.12590	0.37917
TriHeadAttention	0.02397	0.18373
Baseline	0.04427	0.06439

Table 4: Comparative performance analysis on test set for Subtask 1.

Team_Name	F1-Score	Precision	Recall
Volta	0.48166	0.50061	0.46409
HOMADOS	0.40737	0.41206	0.40278
WVOQ	0.26787	0.24265	0.29894
CSECUDSG	**0.11983**	**0.07952**	**0.24303**
YNUHPCC	0.09111	0.18583	0.06035
Baseline	0.00952	0.03368	0.00554

Table 5: Comparative performance analysis on test set for Subtask 2.

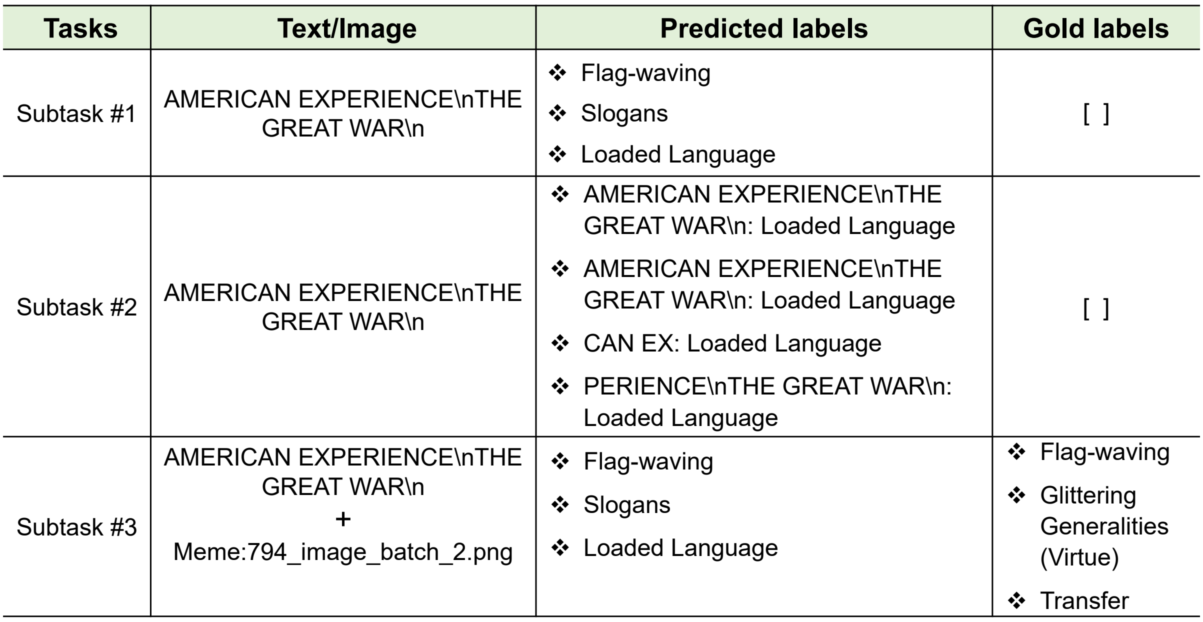
Team_Name	F1-Macro	F1-Micro
Alpha	0.27315	0.58109
MinD	0.24389	0.56623
1213Li	0.22830	0.54860
CSECUDSG	**0.12117**	**0.51312**
LIIR	0.18807	0.49835
WVOQ	0.23957	0.47779
Baseline	0.05152	0.07062

Table 6: Comparative performance analysis on test set for Subtask 3.

3.4 Discussion

In this section, we discuss the contribution of each model's performance against the combined system. For subtask 1, we showed the individual model's performance on the test set in Table 7. From the table, we can see that the decision tree classifier achieved a score of 0.335 where the score is 0.426 and 0.480 in the case of the logistic regression classifier and DistilBERT model respectively. Analyzing this individual model's score, we can say that we achieved the highest score from the DistilBERT model. After applying majority voting, our score increased to 0.008% and the final score is 0.48894 which means that the ensemble of three individual models can detect better than individual models.

In subtask 3, from the Table 8 we observe that the fine-tuned DistilBERT model provides a little better score than the majority voting based model. However, for the multi-modal task, both text and image contexts are important, therefore we consider the majority voting based model.



Tasks	Text/Image	Predicted labels	Gold labels
Subtask #1	AMERICAN EXPERIENCE\nTHE GREAT WAR\n	❖ Flag-waving ❖ Slogans ❖ Loaded Language	[ ]
Subtask #2	AMERICAN EXPERIENCE\nTHE GREAT WAR\n	❖ AMERICAN EXPERIENCE\nTHE GREAT WAR\n: Loaded Language ❖ AMERICAN EXPERIENCE\nTHE GREAT WAR\n: Loaded Language ❖ CAN EX: Loaded Language ❖ PERIENCE\nTHE GREAT WAR\n: Loaded Language	[ ]
Subtask #3	AMERICAN EXPERIENCE\nTHE GREAT WAR\n + Meme:794_image_batch_2.png	❖ Flag-waving ❖ Slogans ❖ Loaded Language	❖ Flag-waving ❖ Glittering Generalities (Virtue) ❖ Transfer

Figure 5: Erroneous detection of persuasive techniques.

Method	F1-Score
CSECUDSG	0.48894
Performance of Individual Models	
−Logistic Regression	0.33585
−Decision Tree	0.42685
−Fine-tuned DistilBERT model	0.48064

Table 7: Individual model's performance for Subtask 1.

Method	F1-Score
CSECUDSG	0.51312
Performance of Individual Models	
−MKCNN+LSTM and MLP model	0.36836
−Fine-tuned ResNet50 model	0.46449
−Fine-tuned DistilBERT model	0.52899

Table 8: Individual model's performance for Subtask 3.

Further, we look into the reason behind the inaccuracy of multiple labels detected by our systems in all the subtasks. For this purpose, we have shown some examples in Figure 5. We noticed that due to the imbalance of labels in the dataset, our systems could not detect the labels which are present in less amount. As the percentage of these three labels i.e. 'Loaded Language', 'Smears', and 'Name calling/Labeling' are higher than the other labels, our system detects these three labels considerably but overlooks other labels.

4 Conclusion and Future Directions

In this paper, we traversed different classification approaches along with a rich set of transfer learning features to tackle the challenges of the task. To predict the multiple labels in subtask 1 and 3, we exploited a unified architecture based on three different models. However, for span and technique classification in subtask 2, we used the pre-trained BERT model where the SemEval-2020 task 11 dataset is used to ameliorate the performance.

In the future, for subtask 1 and subtask 2, we have a plan to employ more pre-processing techniques and to conduct our experiment on efficient classifiers. Besides, we want to use deep learning models as well as other transfer learning fine-tuned models i.e. RoBERTa, BERT, GPT. For subtask 3, we plan to incorporate various image datasets to get more efficacious features.

References

Sami Abu-El-Haija, Nisarg Kothari, Joonseok Lee, Paul Natsev, George Toderici, Balakrishnan Varadarajan, and Sudheendra Vijayanarasimhan. 2016. Youtube-8m: A Large-Scale Video Classification Benchmark. *arXiv preprint arXiv:1609.08675*.

Ilias Chalkidis, Manos Fergadiotis, Prodromos Malakasiotis, and Ion Androutsopoulos. 2019. Large-Scale Multi-label Text Classification on EU Legislation. *arXiv preprint arXiv:1906.02192*.

Weiwei Cheng and Eyke Hüllermeier. 2009. Combining Instance-based Learning and Logistic Regression for Multilabel Classification. *Machine Learning*, 76(2-3):211–225.

Giovanni Da San Martino, Alberto Barrón-Cedeño, Henning Wachsmuth, Rostislav Petrov, and Preslav Nakov. 2020. SemEval-2020 Task 11: Detection of Propaganda Techniques in News Articles. In *Proceedings of the Fourteenth Workshop on Semantic Evaluation (SemEval)*, pages 1377–1414, Barcelona (online). International Committee for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL:HLT)*, pages 4171–4186.

Dimiter Dimitrov, Bishr Bin Ali, Shaden Shaar, Firoj Alam, Fabrizio Silvestri, Hamed Firooz, Preslav Nakov, and Giovanni Da San Martino. 2021. Task 6 at SemEval-2021: Detection of Persuasion Techniques in Texts and Images. In *Proceedings of the 15th International Workshop on Semantic Evaluation*, SemEval '21, Bangkok, Thailand.

Hao Fei, Yue Zhang, Yafeng Ren, and Donghong Ji. 2020. Latent Emotion Memory for Multi-Label Emotion Classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34,05, pages 7692–7699.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

Dawid Jurkiewicz, Łukasz Borchmann, Izabela Kosmala, and Filip Graliński. 2020. ApplicaAI at SemEval-2020 Task 11: On RoBERTa-CRF, Span CLS and Whether Self-Training Helps Them. *arXiv preprint arXiv:2005.07934*.

Liunian Harold Li, Mark Yatskar, Da Yin, Cho-Jui Hsieh, and Kai-Wei Chang. 2019. VisualBERT: A Simple and Performant Baseline for Vision and Language. *arXiv preprint arXiv:1908.03557*.

Yikuan Li, Hanyin Wang, and Yuan Luo. 2020. A Comparison of Pre-trained Vision-and-Language Models for Multimodal Representation Learning across Medical Images and Reports. In *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 1999–2004. IEEE.

Weijie Liu, Peng Zhou, Zhe Zhao, Zhiruo Wang, Haotang Deng, and Qi Ju. 2020. FastBERT: a Self-distilling BERT with Adaptive Inference Time. *arXiv preprint arXiv:2004.02178*.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arXiv preprint arXiv:1907.11692*.

Edward Loper and Steven Bird. 2002. NLTK: The Natural Language Toolkit. *arXiv preprint cs/0205028*.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient Estimation of Word Representations in Vector Space. *arXiv preprint arXiv:1301.3781*.

Gaku Morio, Terufumi Morishita, Hiroaki Ozaki, and Toshinori Miyoshi. 2020. Hitachi at SemEval-2020 Task 11: An Empirical Study of Pre-Trained Transformer Family for Propaganda Detection. In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 1739–1748.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research (JMLR)*, 12:2825–2830.

S Rasoul Safavian and David Landgrebe. 1991. A Survey of Decision Tree Classifier Methodology. *IEEE transactions on systems, man, and cybernetics*, 21(3):660–674.

Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. DistilBERT, a Distilled Version of BERT: Smaller, Faster, Cheaper and Lighter. *arXiv preprint arXiv:1910.01108*.

Lingyun Song, Jun Liu, Buyue Qian, Mingxuan Sun, Kuan Yang, Meng Sun, and Samar Abbas. 2018. A Deep Multi-Modal CNN for Multi-Instance Multi-Label Image Classification. *IEEE Transactions on Image Processing*, 27(12):6025–6038.

Shiping Wen, Weiwei Liu, Yin Yang, Pan Zhou, Zhenyuan Guo, Zheng Yan, Yiran Chen, and Tingwen Huang. 2020. Multilabel Image Classification via Feature/Label Co-projection. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*.

Yuxin Wu, Alexander Kirillov, Francisco Massa, Wan-Yen Lo, and Ross Girshick. 2019. Detectron2. `https://github.com/facebookresearch/detectron2`.

Chenyang Zhang, Christine Kaeser-Chen, Grace Vesom, Jennie Choi, Maria Kessler, and Serge Belongie. 2019. The iMet Collection 2019 Challenge Dataset. *arXiv preprint arXiv:1906.00901*.