DEEP LEARNING FOR ARABIC COMPUTATIONAL LINGUISTICS

CLASP centre for linguistic theory and studies in probability

Kathrein Abu Kwaik

AGENDA

- Deep learning for Arabic NLP
- Deep learning for sentiment analysis
- Challenges
- Our Works /Expirements
- Reproduce Results
- Sentiment Analysis with BERT

DEEP LEARNING FOR ARABIC NLP

1. **Caption Generation**
2. **Language modelling**
3. **ATM**
4. **Dialect Detection**
5. **Text Categorization**
6. Sentiment Analysis
7. Question Answering
8. Automatic Diacritization
9. OCR
10. ASR

CAPTION GENERATION

- Unfortunately, it is almost untouched in the ANLP community.
- Researchers from Dallas university, Texas addressed the problem.
- The results were very encouraging as they represent the first approach for Arabic caption generation. Moreover, they were much higher than the simple approach of generating English captions and automatically translating them into Arabic.

Approach	BLEU-1 Score
English	48.4
English-Arabic (Google Translate)	27.2
Our Approach	**34.8**

CAPTION GENERATION

Object Extraction from Images as Input

⬇

Stage I : Mapping Image Fragments to Arabic Root Words

⬇

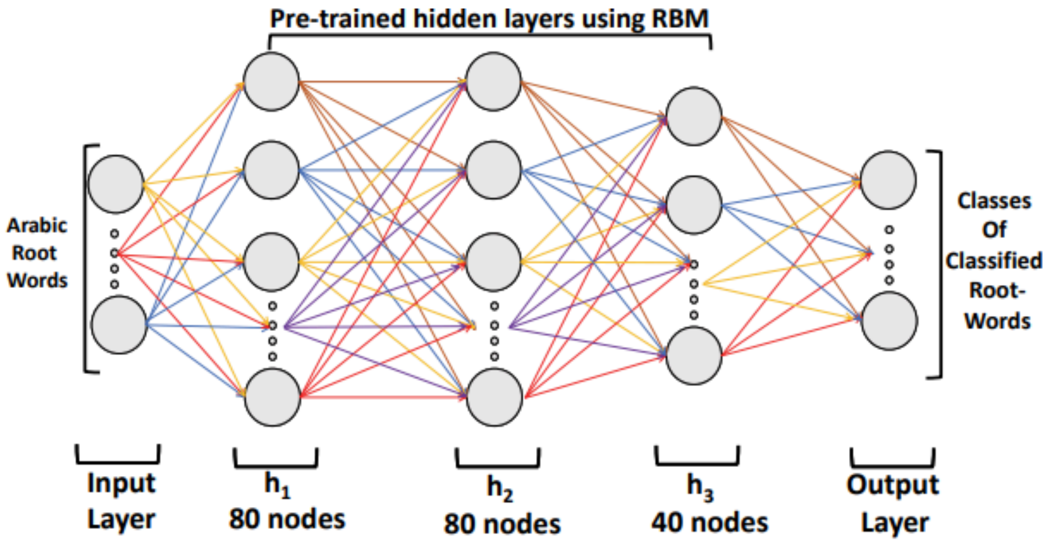
Stage II : Adding Vowels and Consonants to Root Words using Deep Belief Networks

⬇

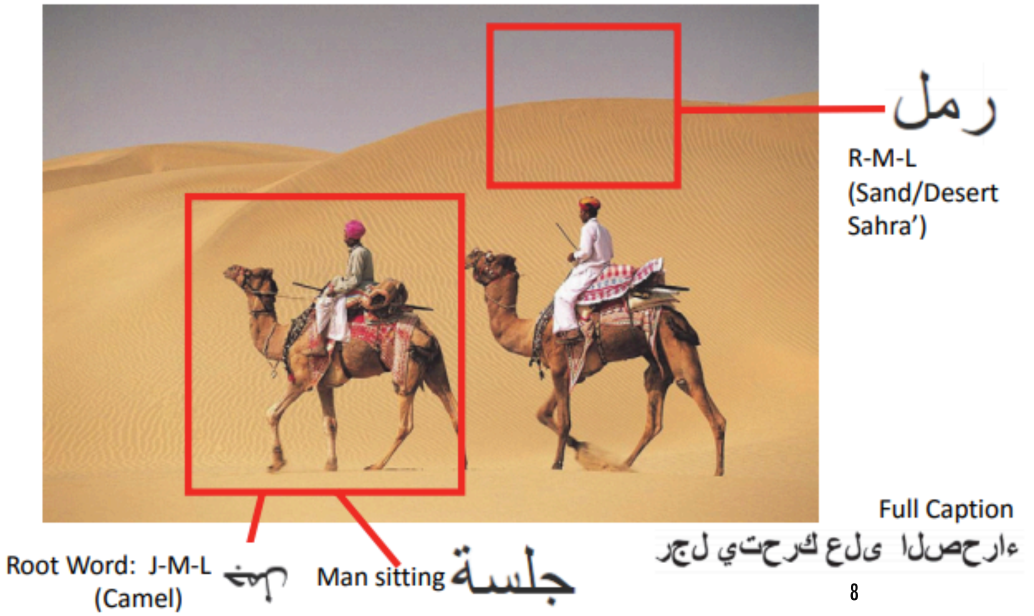
Stage III : Leveraging dependency trees to rank most appropriate sentences from extracted images

⬇

Image captions directly in Arabic as Output

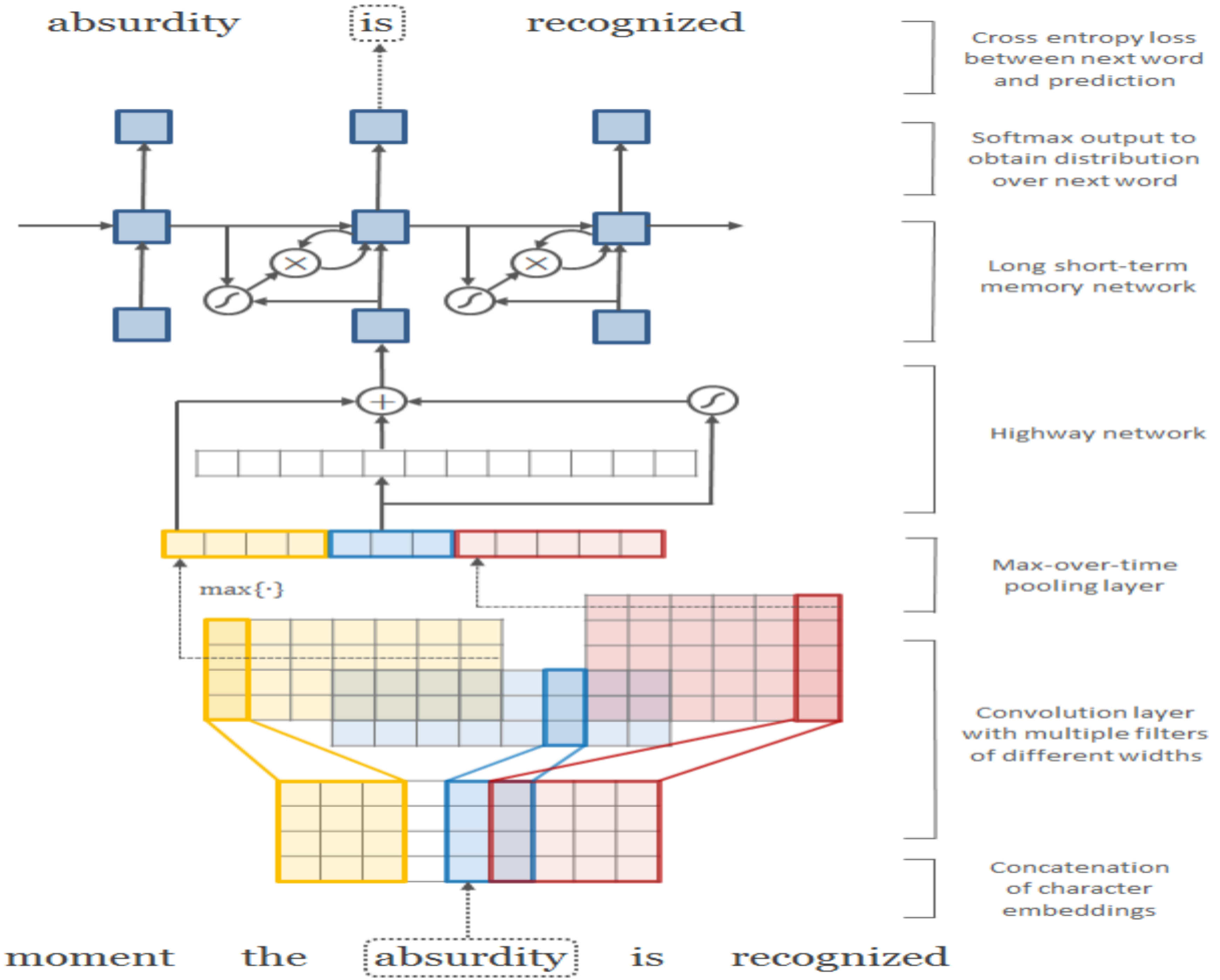


Deep Belief Network to add vowels to root words



LANGUAGE MODELING

- Researchers from Harvard and NY universities proposed a character-level LM that can work on English as well as other languages such as Arabic. The proposed model applies CNN on input characters before feeding them into LSTM RNN-LM.
- The results for the Arabic language showed that the proposed LM outperformed various baselines working on word level or morpheme level.


absurdity
is
recognized
Cross entropy loss between next word and prediction
Softmax output to obtain distribution over next word
Long short-term memory network
Highway network
Max-over-time pooling layer
max{·}
Convolution layer with multiple filters of different widths
Concatenation of character embeddings
moment
the
absurdity
is
recognized

AUTOMATIC MACHINE TRANSLATION (ATM)

- Researchers proposed to address the Arabic-English machine transliteration problem using DBN, which contains multiple layers of restricted Boltzmann machines RBM.
- The proposed approach has three important parts. The first one is the source encoder, which deals with source words by converting them to dimensional binary vectors, then feeding them into first layer in the source encoder, the output of each layer is considered as an input to the next layer.
- The sec- ond part called joint layer. This layer uses the output of the source encoder as an input in order to get a state of hidden neurons, and infer an output state to use as input to the top level of the out- put encoder.
- The third part is the target encoder. Within this part, the output vector is decoded by traversing down words through the output encoder.

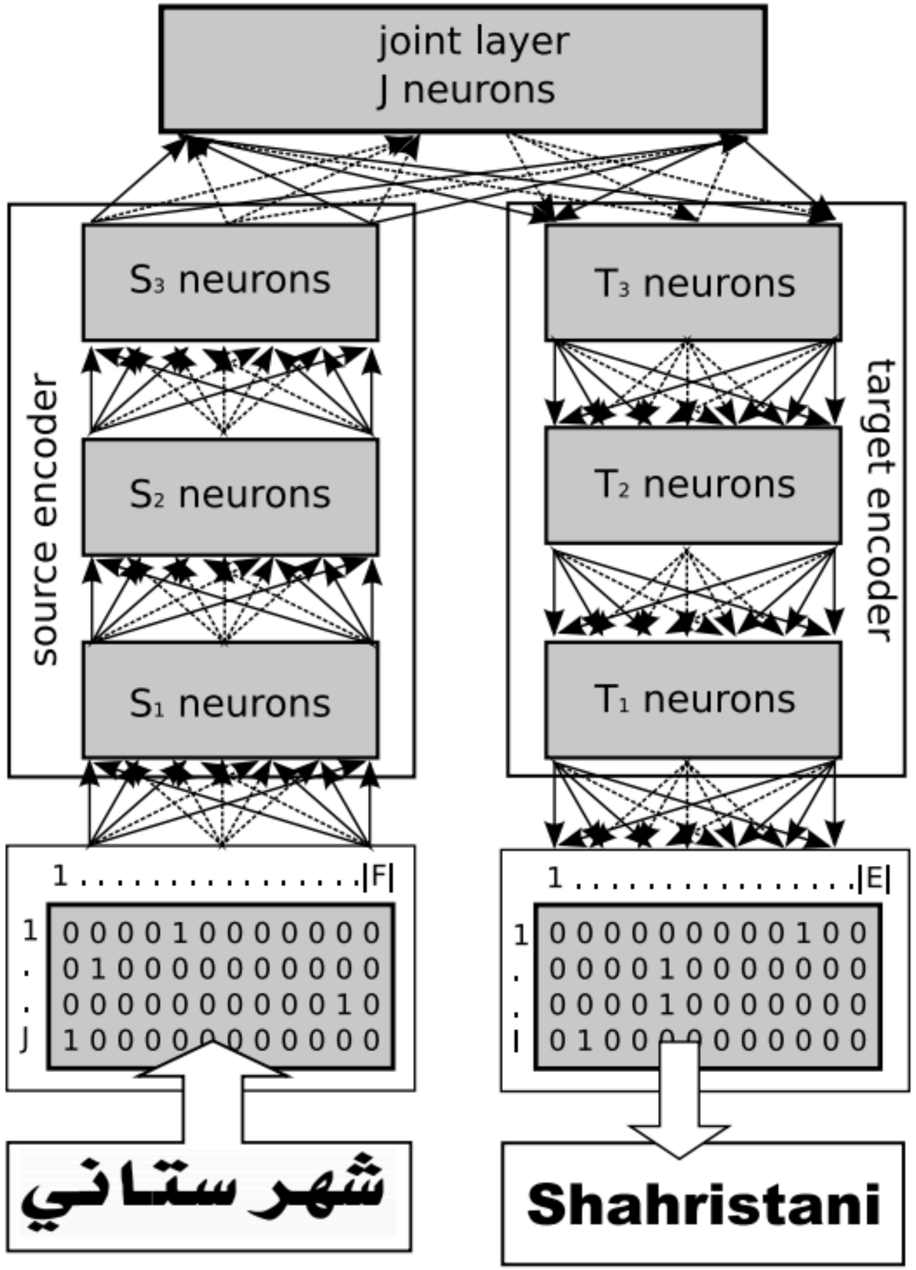


Figure 1: A schematic representation of our DBN for transliteration.

number of nodes				CER [%]		
S_1,T_1	S_2,T_2	S_3,T_3	J	train	dev	eval
400	500	600	1800	0.3	27.2	28.1
400	400	400	1200	0.7	26.1	25.2
400	350	300	900	1.8	25.1	24.3
400	350	300	1000	1.7	24.8	24.0
400	**350**	**300**	**1500**	**1.3**	**24.1**	**22.7**
400	350	300	2000	0.2	24.2	23.5

DIALECT DETECTION

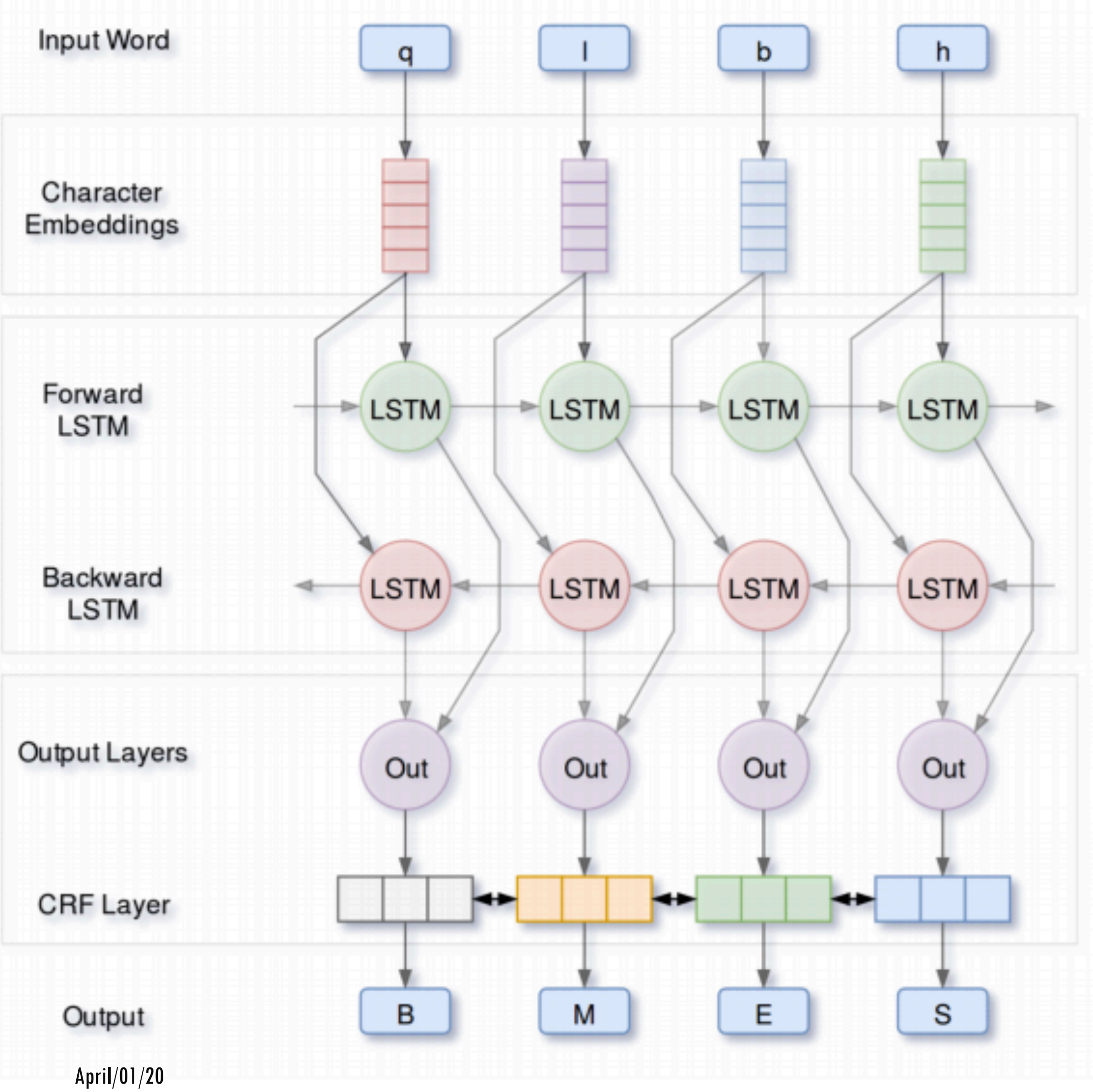
- Researchers describe their character-level NN for the Arabic dialects identification task of the DSL challenge .
- Given a sequence of characters, their model embeds each character in vector space, runs the sequence through multiple convolutions with different filter widths, and pools the convolutional representations to obtain a hidden vector representation of the text that is used for predicting the language or dialect.
- The implementation of their approach is publicly available 15 and the obtained F-measure is 48.3%

DIALECT DETECTION

The neural network has the following structure:

- Input layer: mapping the character sequence c to a vector sequence x. The embedding layer is followed by dropout.
- Convolutional layers: multiple parallel convolutional layers, mapping the vector sequence x to a hidden sequence h. Each convolution is followed by a Rectified Linear Unit (ReLU) nonlinearity The outputs of all the convolutional layers are concatenated.
- Pooling layer: a max-over-time pooling layer, mapping the vector sequence h to a single hidden vector h representing the entire sequence.
- Fully-connected layer: one hidden layer with a ReLU non-linearity and dropout, mapping h to the final vector representation of the text, h 0 .
- Output layer: a softmax layer, mapping h 0 to a probability distribution over labels l.

DIALECTAL ARABIC SEGMENTATION



The usage of a character-level BLSTM network combined with the conditional random field (CRF) algorithm to build a segmenter for the Egyptian dialect

DEEP LEARNING FOR SENTIMENT ANALYSIS

Different Classification:

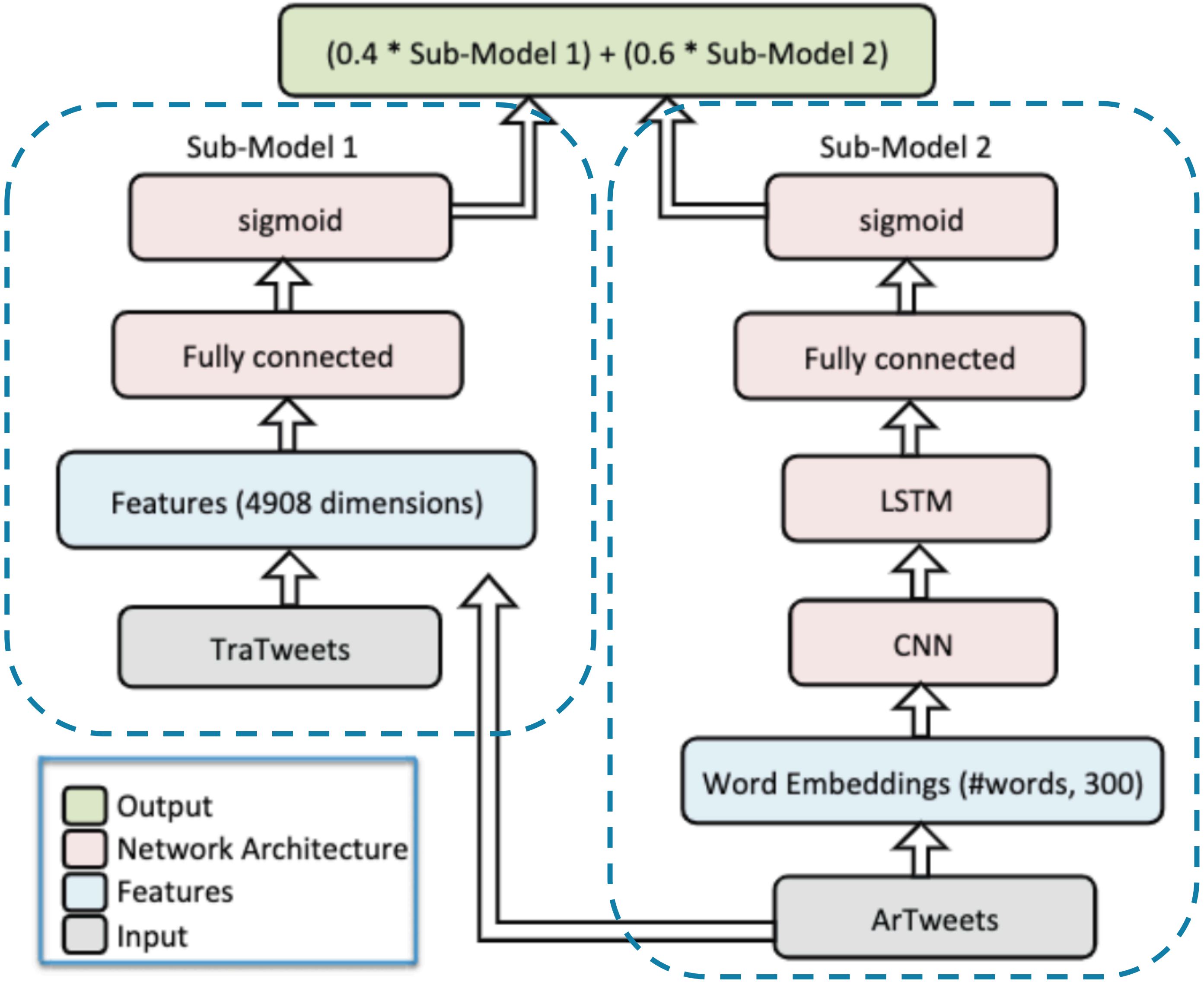
1. Intensity of Classification (joy, fear, sadness, anger)
2. Polarity (positive, negative, mix, neutral)
3. Degree of Polarity (very negative, negative, neutral, positive, very positive)

1. SEDAT: SENTIMENT AND EMOTION DETECTION IN ARABIC TWEETS

- Detect and predict the intensity of sentiment and emotions in Arabic Tweets
- Features: word embeddings + semantic features (English)
- Out put:
  - Emotion (No, low, moderate, high)
  - Intensity (most positive --- most negative)
  - Sentiment(real value from -1 to +1)

2. SEDAT

English + Arabic

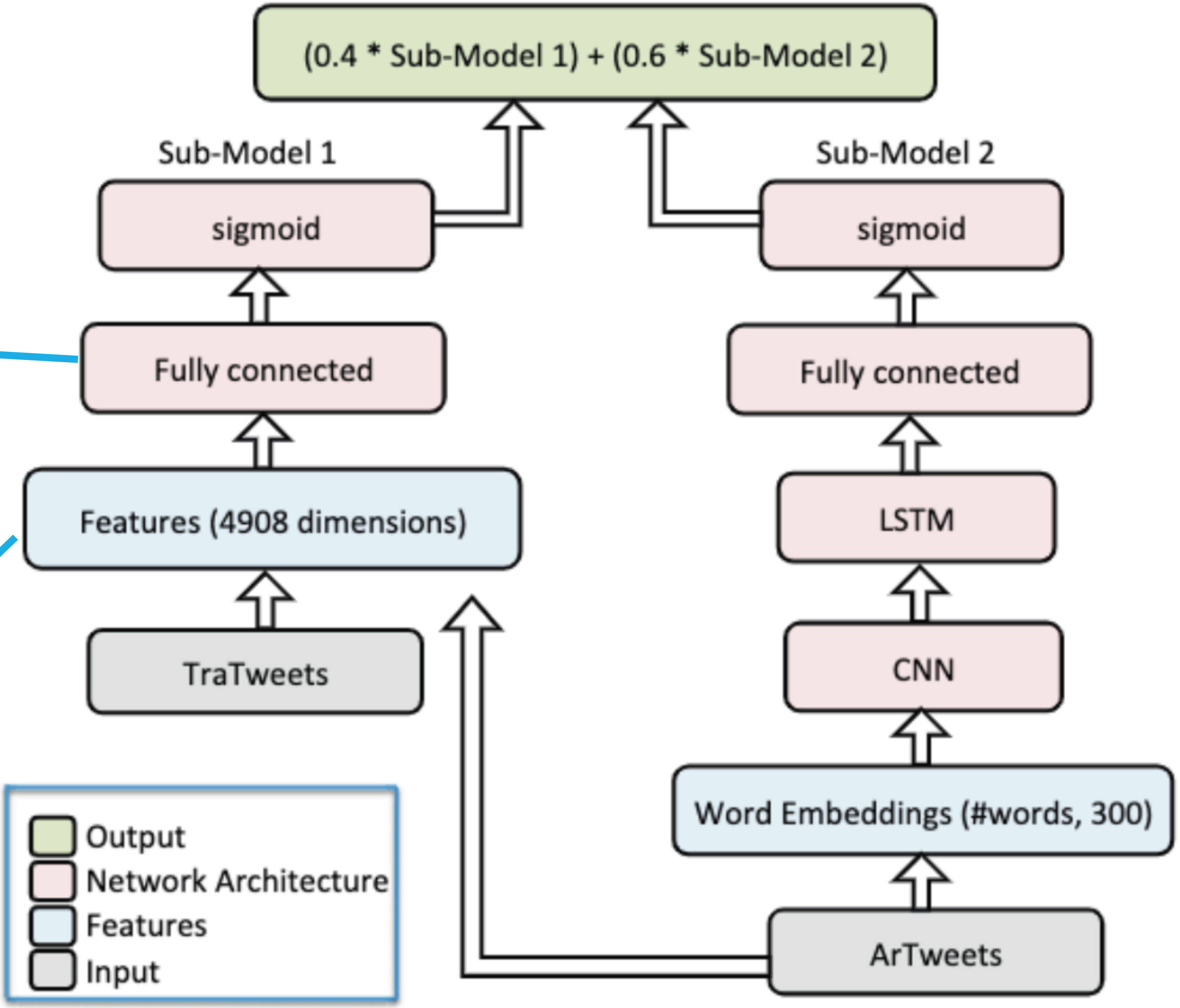


Arabic

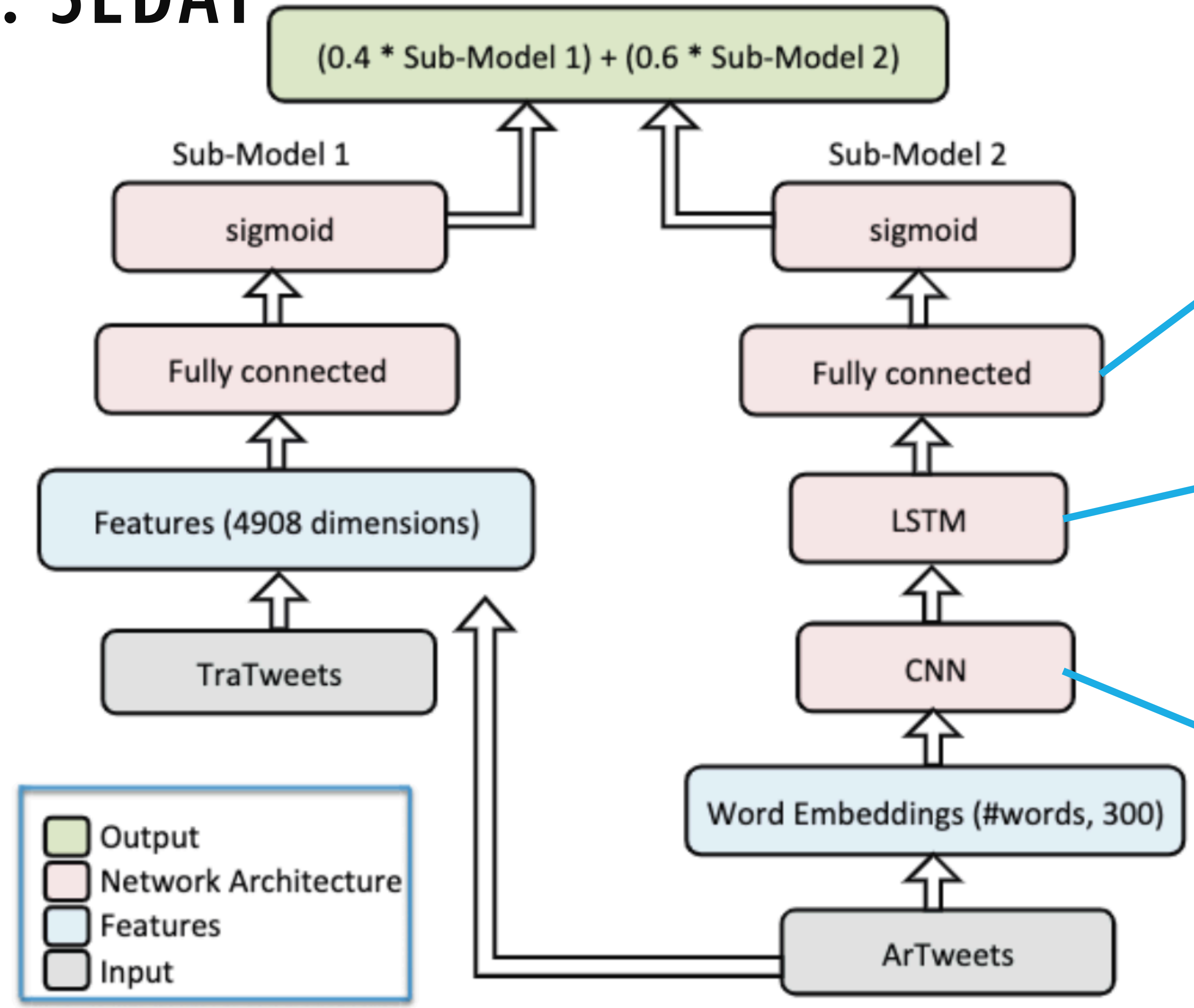
2. SEDAT

- 3-dense layer (500,200,80)
- RELU act.fun.

- AffectiveTweets-142
- Doc2Vec-600
- Arabic Feature-5
- DeepEmoji-64
- UnsupervisedLearning-4096
- EmojiFeature-1



2. SEDAT



- 2-dense layer (200,80)
- RELU act.fun.

- 256 neuron
- Drop out 0.3

- 64 filter
- 3 kernel size
- RELU
- Max-pool size 2

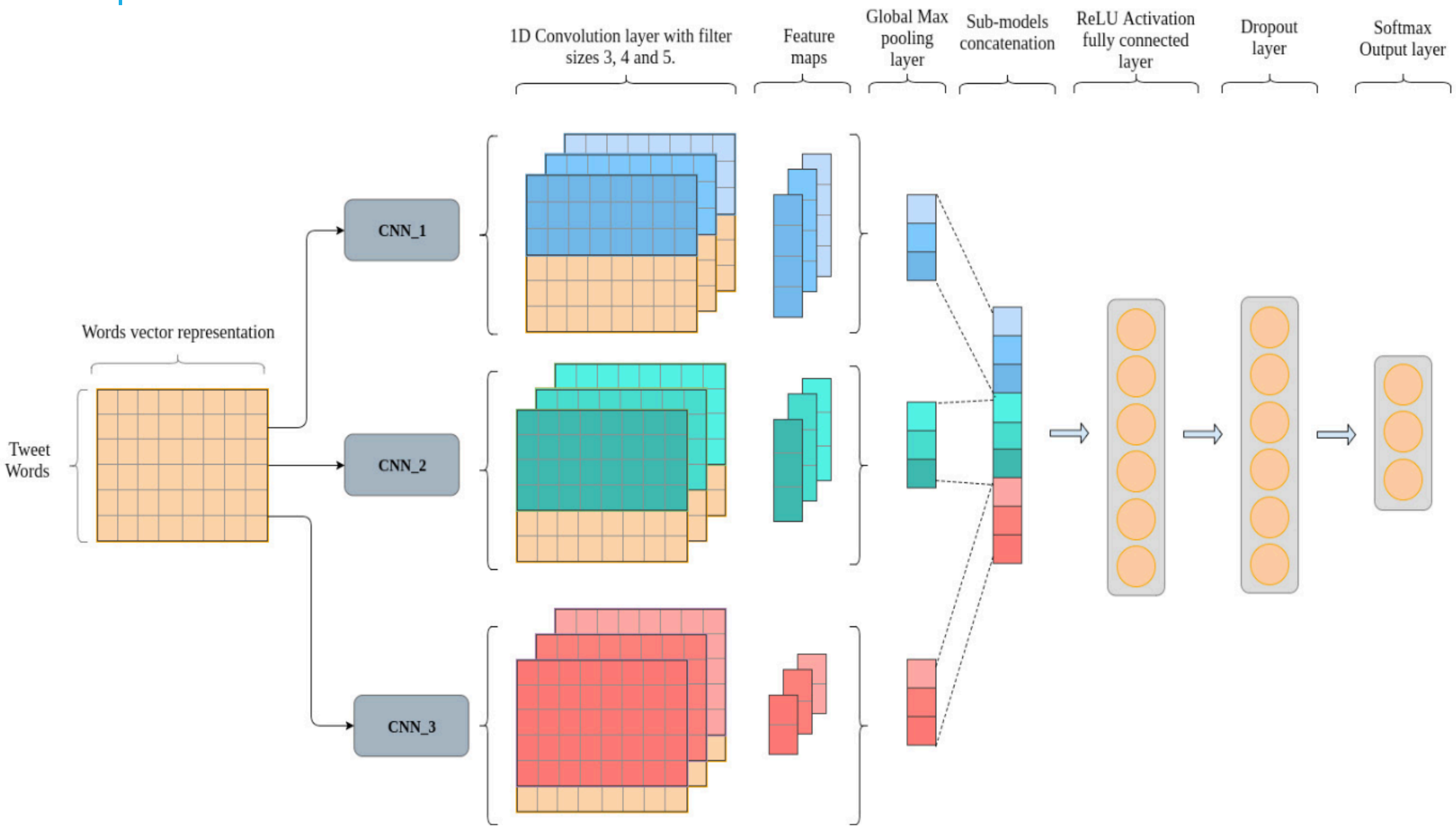
2. SEDAT

Analysing the spearman correlation scores of SEDAT system for each dialect

2. SENTIMENT ANALYSIS OF ARABIC TWEETS

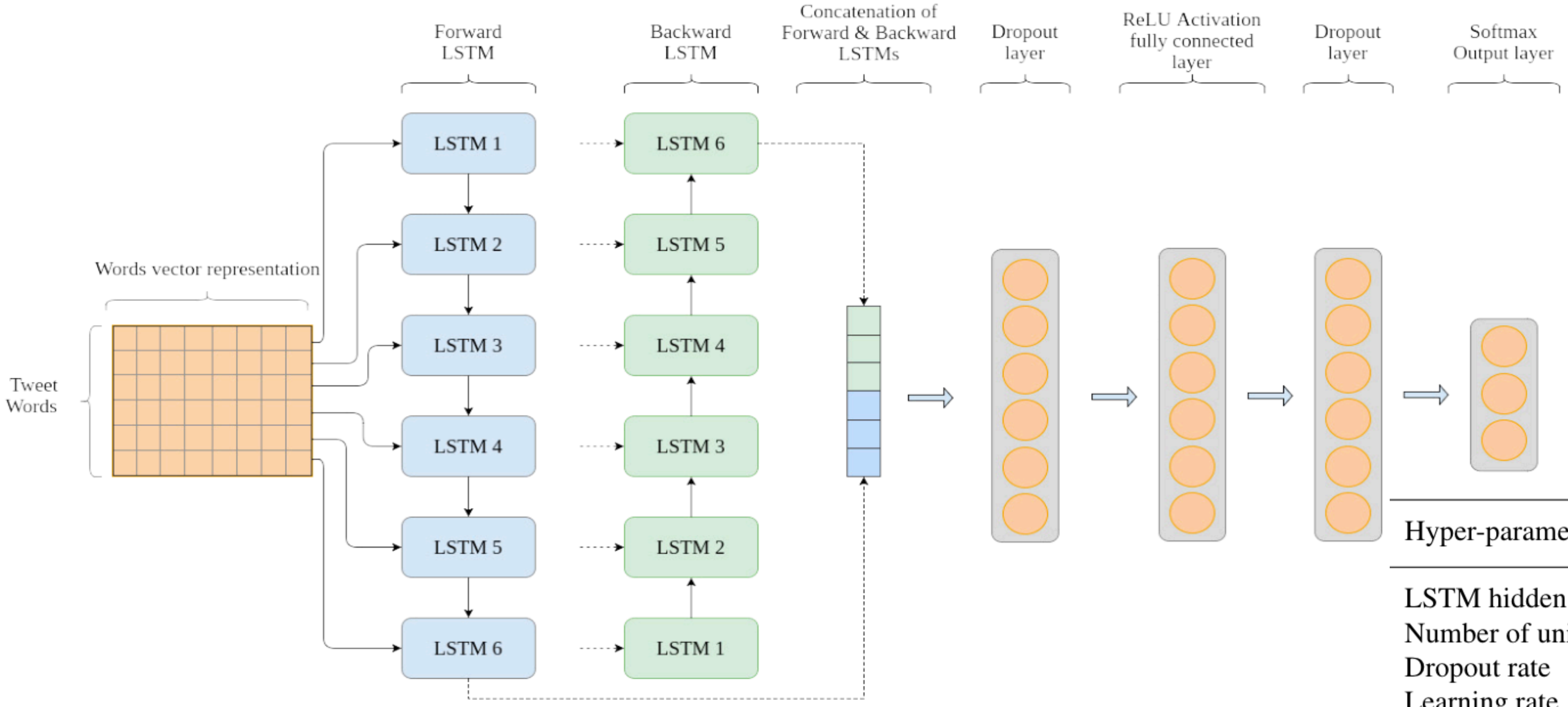
- an ensemble model, combining Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) models, to predict the sentiment of Arabic tweets.
- achieves an F1-score of 64.46, on the Arabic Sentiment Tweets Dataset (ASTD)

2. SENTIMENT ANALYSIS OF ARABIC TWEETS



Hyper-parameter	Value
Filter sizes	[3, 4, 5]
Number of filters	200
Number of units in fully connected layer	30
Dropout rate	0.5
Learning rate	0.001
Number of epochs	10
Batch size	50

2. SENTIMENT ANALYSIS OF ARABIC TWEETS



Hyper-parameter	Value
LSTM hidden state dimension	200
Number of units in fully connected layer	30
Dropout rate	0.5
Learning rate	0.001
Number of epochs	10
Batch size	50

2. SENTIMENT ANALYSIS OF ARABIC TWEETS

Model	Accuracy (%)	F1-score (%)
CNN (fully connected layer size=100)	64.30	64.09
LSTM (dropout rate=0.2)	64.75	62.08
Ensemble model	**65.05**	**64.46**
Previous best model (RNTN)	58.5	53.6

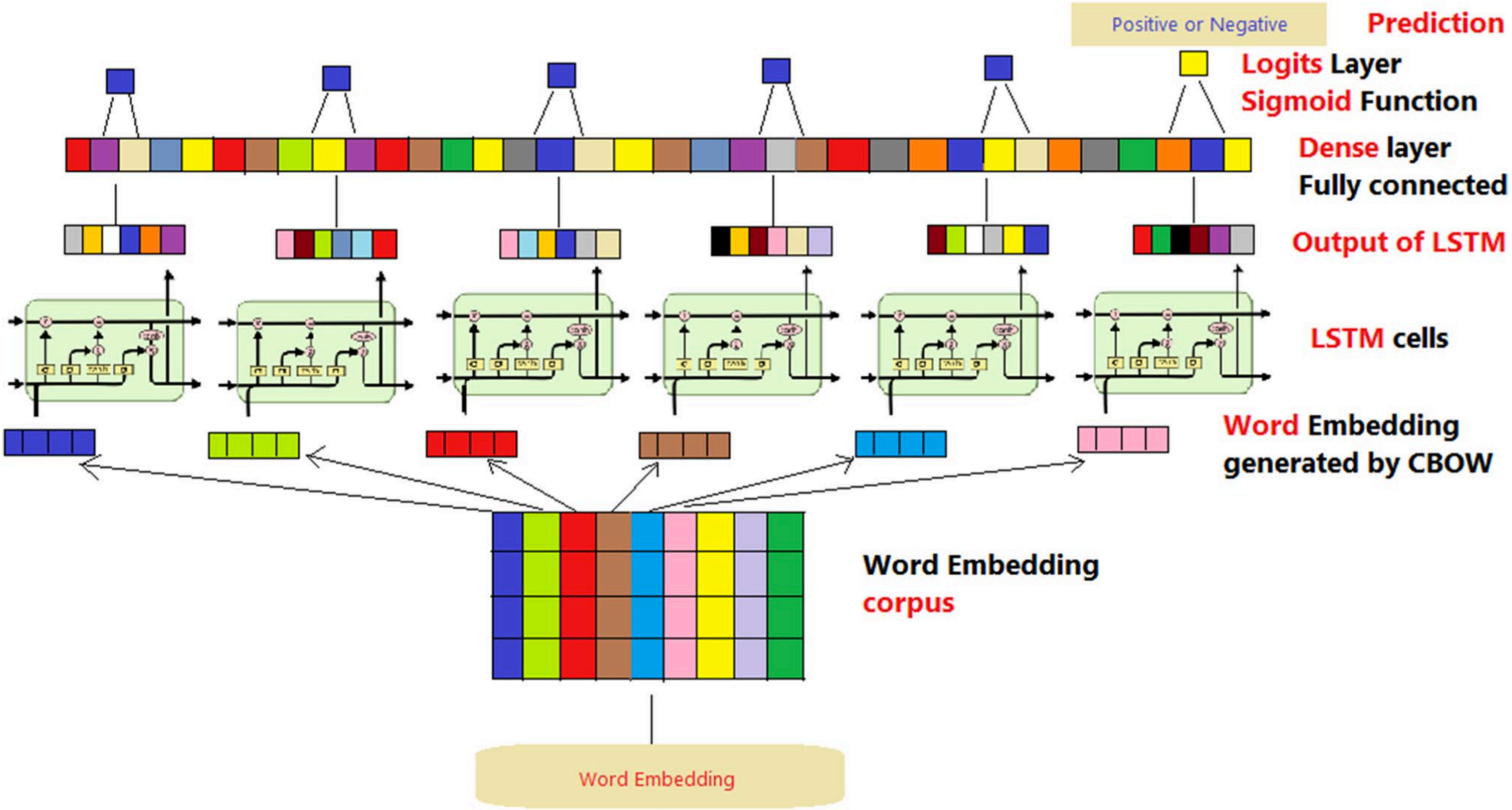
3. DEEP LEARNING APPROACH FOR ARABIC SA

- Introduce a corpus of 40k labeled Arabic tweets spanning several topics.
- Present three deep learning models, namely CNN, LSTM and RCNN, for Arabic sentiment analysis.
- Validate the performance of the three models on the proposed corpus. The experimental results indicate that LSTM with an average accuracy of 81.31% outperforms CNN and RCNN.

3. DEEP LEARNING APPROACH FOR ARABIC SA

Twitter corpus statistics

Total number of tweets	40,000
Number of positive tweets	20,000
Number of negative tweets	20,000
Number of words	359,818
Max tweet token	39
Number of tokens	1,953,869
Average tokens per tweet	17



Layers of LSTM

3. DEEP LEARNING APPROACH FOR ARABIC SA

Average performance measures for various splits with LSTM model

Model	Split	AVG accuracy (%)	AVG recall (%)	AVG precision (%)	AVG f-score(%)
LSTM	(80%, 20%)	81.49	**81.6**	81.8	**81.49**
	(70%, 30%)	**81.53**	80.95	**82.31**	81.43
	(60%, 40%)	80.91	80.21	81.86	80.84
Total AVG		81.31	80.9	81.99	81.25

CHALLENGES

1. Complex Morphology
2. Dialectal Arabic
3. Arabizi (Romanized Arabic)
4. >100 forms of Arabic Alphabets
5. Limited Resource
6. Social media text: spelling inconsistencies, abb., Slang, repeat letters for exaggeration, lack of capitalization, Ironic expression.
7. Some Arabic names are sentiment adj.
8. Same root with different sentiment (Discrimination, Excellent) (تمييز وامتياز) (ميز)

OUR WORKS

1. Can Modern Standard Arabic Approaches be used for Arabic Dialects?
2. Build DL model to predict SA of Dialectal Arabic
3. Apply Transfer learning and weak supervision to build DASA corpus
4. Results Reproducibility
5. Apply BERT

1. CAN MODERN STANDARD ARABIC APPROACHES BE USED FOR ARABIC DIALECTS?

- Build the first Levantine corpus for Sentiment Analysis (SA)
- Investigate the usage of off-the-shelf models that have been built for Modern Standard Arabic (MSA) on this corpus of Dialectal Arabic (DA).
- apply the models on DA data, showing that their accuracy does not exceed 60%.
- Build our own models involving different feature combinations and machine learning methods for both MSA and DA and achieve an accuracy of 83% and 75% respectively

1. CAN MODERN STANDARD ARABIC APPROACHES BE USED FOR ARABIC DIALECTS?

- Building Shami-Senti
  - Automatic annotation: using different lexicon , compare with human annotation over 1000 sample, Agreement < 80%
  - Manual Annotation: two native speakers, over 2000 sample, $\kappa = 0.838$

Corpus	NEG	POS	Mix
Shami-Senti	935	1064	243
LARB 3 Balanced	6580	6578	6580
LABR 2 Balanced	6578	6580	
ASTD	1496	665	738

The number of instances per category in Shami-Senti and other sentiment corpora used in our experiments

1. CAN MODERN STANDARD ARABIC APPROACHES BE USED FOR ARABIC DIALECTS?

- In all experiments, we use the same machine learning algorithms that have been used by the LABR baseline.
- These are:
  1. Logistic Regression (LR)
  2. Passive Aggressive (PA)
  3. Linear Support Vector classifier (LinearSVC)
  4. Naive-Bayes (BNB, MNB, CNB)
  5. Stochastic Gradient Descent (SGD)

3-WAY SENTIMENT CLASSIFICATION (MODEL 2)

Classifier	Accuracy
Ridge Classifier	43
Logistic Regression	46
Passive Aggressive	43
Linear SVC	45
SGD Classifier	50
Multinomial NB	40
Bernoulli NB	44
Complement NB	42

Accuracy of the proposed model trained on LABR3 and tested on Shami-Senti

Trained On MSA → TEST ON DA

Classifier	Accuracy
Ridge Classifier	69
Logistic Regression	67
Passive Aggressive	68
Linear SVC	69
SGD Classifier	68
Multinomial NB	71
Bernoulli NB	71
Complement NB	71

Accuracy of the proposed model 3-class classification trained and tested on Shami-Senti

Trained On DA → TEST ON DA

2-WAY SENTIMENT CLASSIFICATION

	counting 2g		TF_wg 1+2		OUR Model	
Classifier	**LABR**	**Shami**	**LABR**	**Shami**	**LABR**	**Shami**
Ridge Classifier	78	53	81	54	83	57
Logistic Regression	80	57	80	56	82	58
Passive Aggressive	78	53	81	53	82	56
Linear SVC	78	55	81	55	83	58
SGD Classifier	80	53	82	54	83	56
Multinomial NB	78	52	80	53	82	55
Bernoulli NB	76	48	76	47	74	48
Complement NB	78	51	80	53	82	55

Accuracy for binary classifiers with different feature sets trained on the LABR2 dataset and tested on LABR2 and Shami-Senti

Trained on MSA

- Test on MSA
- Test on DA

2-WAY SENTIMENT CLASSIFICATION

Classifier	2 classes
Ridge Classifier	73
Logistic Regression	74
Passive Aggressive	73
Linear SVC	73
SGD Classifier	73
Multinomial NB	74
Bernoulli NB	72
Complement NB	75

Accuracy of the proposed model on binary classification trained and tested on Shami-Senti

Trained On DA → Test on DA

1. CAN MODERN STANDARD ARABIC APPROACHES BE USED FOR ARABIC DIALECTS?

~~Trained On MSA~~ → ~~TEST ON DA~~

Trained On DA → Test on DA

Trained On MSA → Test on MSA

Conclusion

2. LSTM-CNN DEEP LEARNING MODEL FOR SENTIMENT ANALYSIS OF DIALECTAL ARABIC

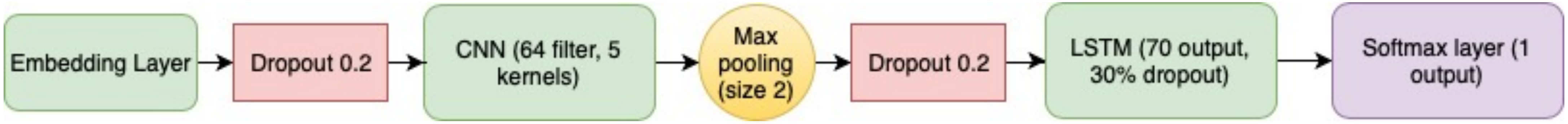
- Investigate the use of Deep Learning (DL) methods for Dialectal Arabic Sentiment Analysis.
- propose a DL model that combines long-short term memory (LSTM) with convolu- tional neural networks (CNN).
- The model achieves an accuracy be- tween 81% binary classification and 66% to 76% accuracy for three-way classification.

2. LSTM-CNN DEEP LEARNING MODEL FOR SENTIMENT ANALYSIS OF DIALECTAL ARABIC

The number of instances per category in the corpora used in our experiments

Corpus	NEG	POS	Neutral
Shami-Senti	935	1,064	243
LABR 3 Balanced	6,580	6,578	6,580
LABR 2 Balanced	6,578	6,580	
LABR 2 Un-Balanced	8,222	42,832	
ASTD	1,496	665	738

KAGGLE EXPIREMENT



KAGGLE EXPIREMENT

Accuracy of the Kaggle model on three-way and binary sentiment classification

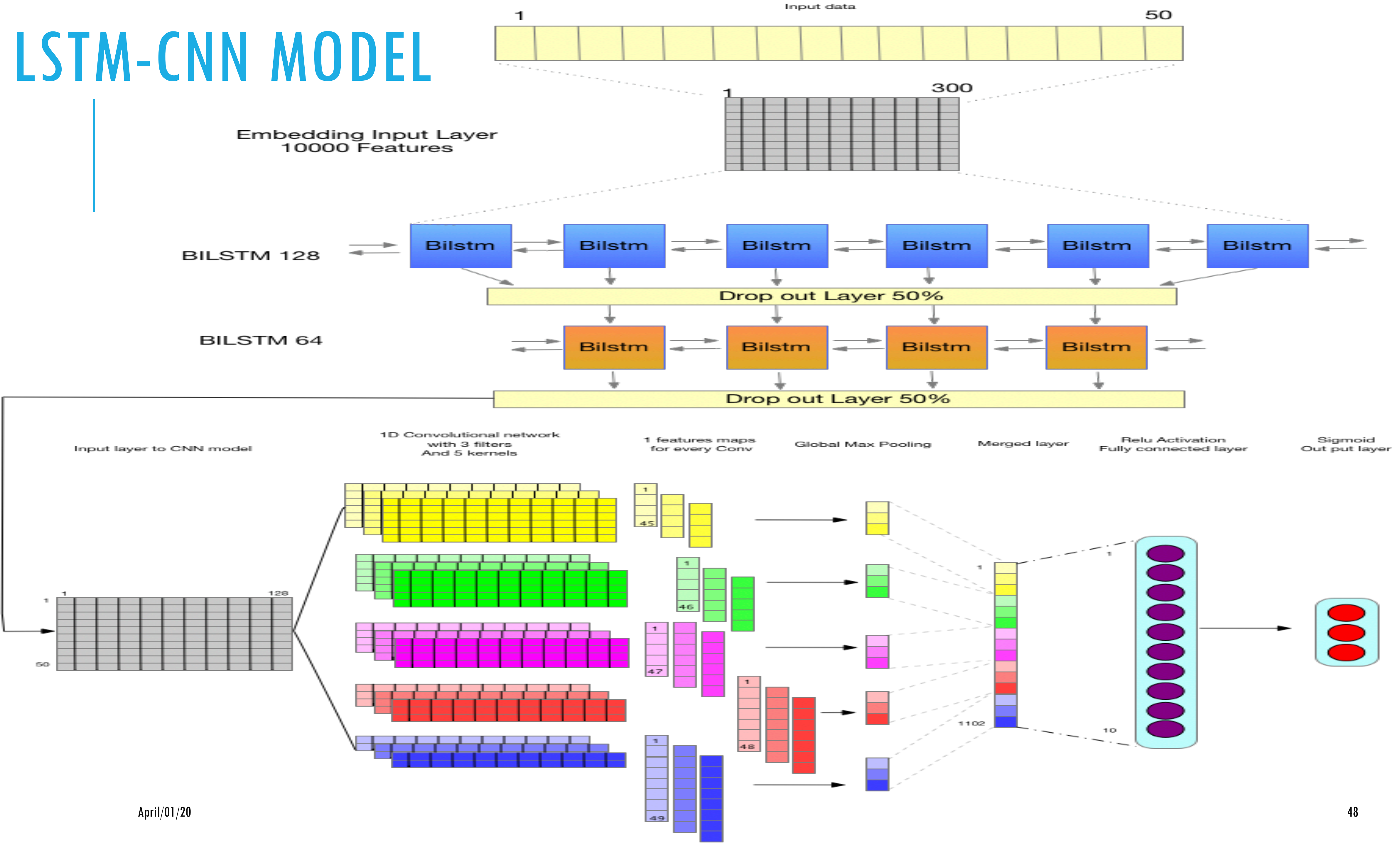
Corpus	Three-way Classification	Binary Classification
Shami-Senti	49%	52.3%
LABR 2 unbalanced		**80.6%**
LABR 2 balanced		53.1%
LABR 3	60%	
ASTD	59.3%	**70.7%**

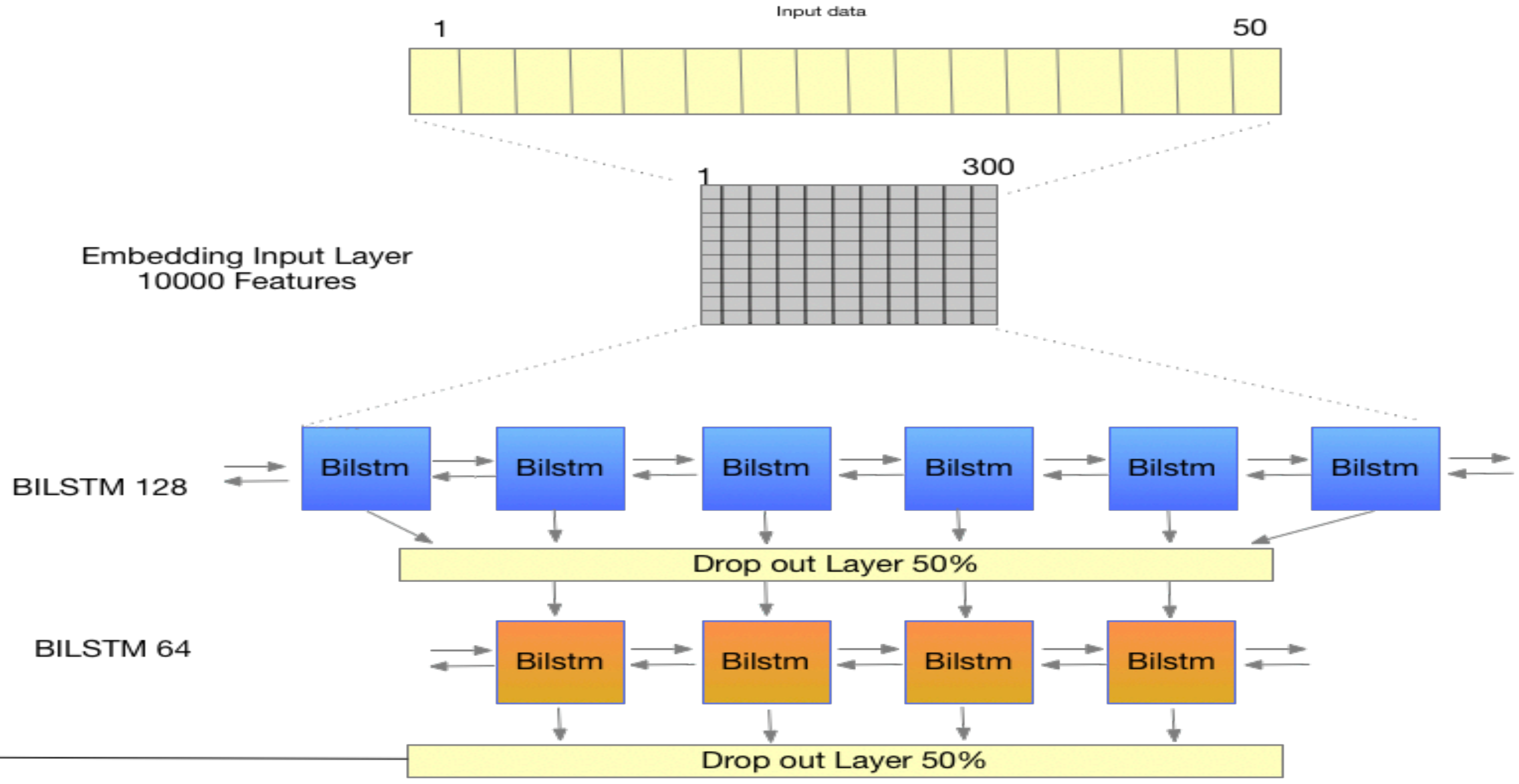
Confusion matrix for the Kaggle model on the ASTD and LABR 2 unbalanced corpora.

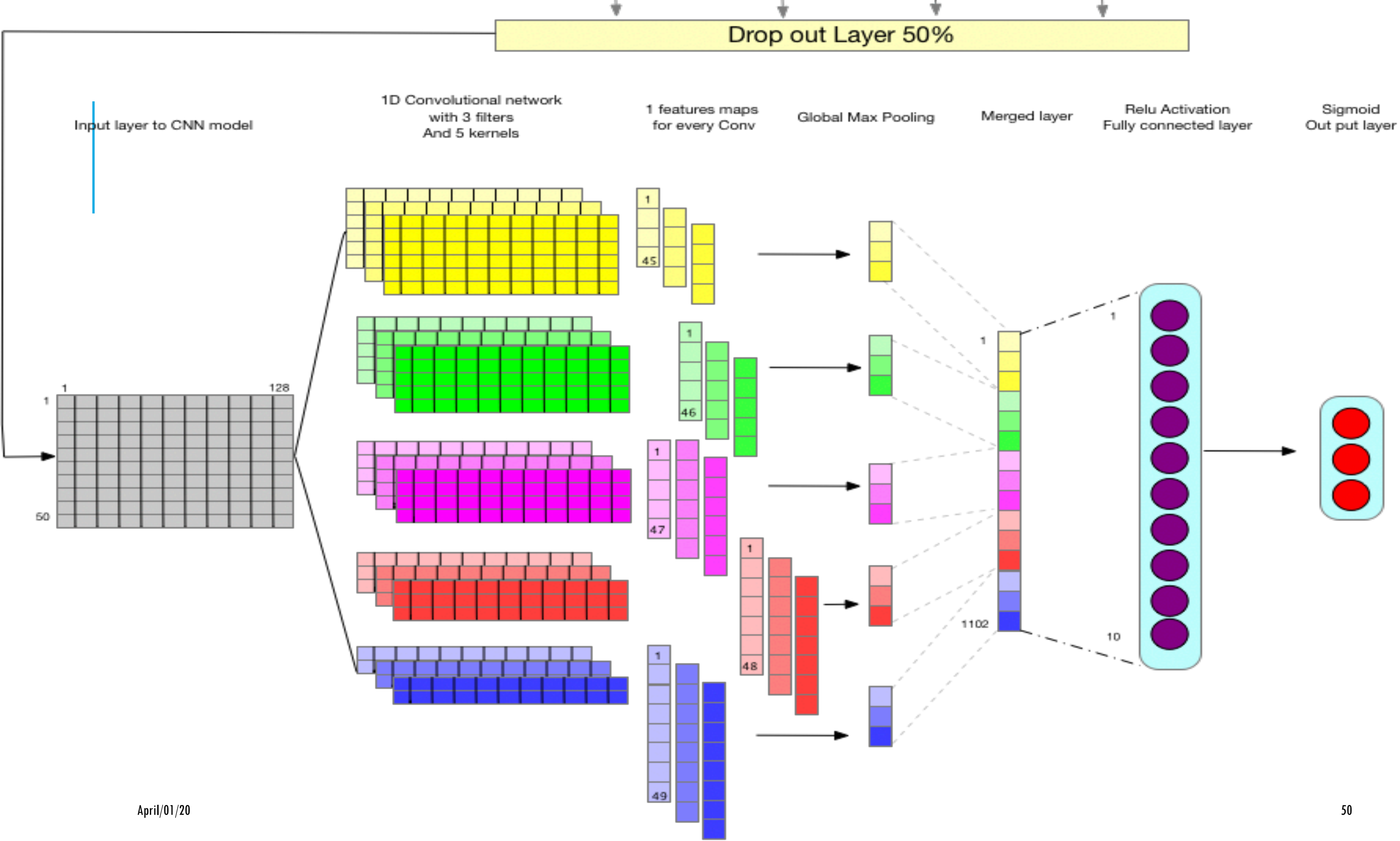
ASTD corpus			
		Predicted	
		Positive	Negative
Actual	Positive	5	**51**
	Negative	12	**147**

LABR 2 unbalanced			
		Predicted	
		Positive	Negative
Actual	Positive	**8153**	387
	Negative	**1591**	78

LSTM-CNN MODEL




Input data
1
50
1
300
Embedding Input Layer
10000 Features
BILSTM 128
Bilstm
Bilstm
Bilstm
Bilstm
Bilstm
Bilstm
Drop out Layer 50%
BILSTM 64
Bilstm
Bilstm
Bilstm
Bilstm
Drop out Layer 50%



Drop out Layer 50%
Input layer to CNN model
1D Convolutional network
with 3 filters
And 5 kernels
1 features maps
for every Conv
Global Max Pooling
Merged layer
Relu Activation
Fully connected layer
Sigmoid
Out put layer
1
128
1
50
1
45
1
46
1
47
1
48
1
49
1
1102
1
10

LSTM-CNN MODEL

Accuracy of the proposed model In addition to the comparing results from the two baselines on the three-way and binary sentiment classification

Corpus	**Three-way Classification**			**Binary Classification**		
	Our Model	Kaggle	LSTM	**Our Model**	Kaggle	LSTM
Shami-Senti	76.4%	49%	53%	93.5%	25.3%	54.5%
LABR 2 unbalanced				80.2%	80.6%	55.34%
LABR 2 balanced				81.14%	53.1%	81%
LABR 3	66.42%	60%	41.9%			
ASTD	68.62%	59.3%	53%	85.58%	70.7%	68.5%

LSTM-CNN MODEL

. Confusion matrix for the proposed model in the ASTD, Shami-Senti and the LABR 2 balanced corpora.

ASTD corpus			
		Predicted	
		Pos	Neg
Actual	Pos	46	18
	Neg	13	138

Shami-Senti			
		Predicted	
		Pos	Neg
Actual	Pos	94	4
	Neg	9	93

LABR2 Balanced			
		Predicted	
		Pos	Neg
Actual	Pos	561	80
	Neg	168	506

3. AN ARABIC TWEETS SENTIMENT ANALYSIS DATASET (ATSAD) USING DISTANT SUPERVISION AND SELF TRAINING

1. Build an Arabic Sentiment Analysis Corpus collected from Twitter, which contains 36K tweets labelled into positive and negative.
2. We employed distant supervision and self-training approaches into the corpus to annotate it.
3. Besides, we release an 8K tweets manually annotated as a gold standard.
4. We evaluated the corpus intrinsically by comparing it to human classification and pre-trained sentiment analysis models.
5. Moreover, we apply extrinsic evaluation methods exploiting sentiment analysis task and achieve an accuracy of 86%.

3.1. BUILD AN ARABIC SENTIMENT ANALYSIS CORPUS (ATSAD)

1. we first build a sentiment emoji lexicon
2. The Lexicon is composed of 91 negative emojis and 306 positive emojis
3. we exploit the emojis and their assigned sentiment and condi- tion the tweet language set to Arabic.
4. We extracted 59k of the tweets using the Twitter API in April 2019.
5. The corpus contains multiple dialects from all over the Arab world.
6. We use the emojis as a noisy (weak) label. EX: If the tweet is fetched by the positive emojis from the lexicon like ☺ then it is labelled as positive.

3.1. ATSAD

	Positive	Negative	Total	Vocabs	Words
Before	30,607	29,232	59,839	95,538	76,2673
After	18,173	18,695	$\boxed{36,868}$	95,057	41,8857

Statistics of the Twitter sentiment analysis corpus (ATSAD) before and after the pre-processing

3.2. EVALUATION

Sample %	Samples	#errors	Accuracy
1%	360	106	70.5%
2%	720	200	72.2%
3%	1,080	293	72.9%
4%	1,400	370	74.3%
5%	1,800	450	75%
10%	3,608	823	77.2%

Human annotation accuracy compared to the emojis based annotation. The first two columns show the percentage and number of the sampled tweets, #_error shows the number of mismatched samples and the Accuracy column calculates the percentage of the matches between both annotations.

3.3. SELF TRAINING ON DISTANT SUPERVISION CORPUS

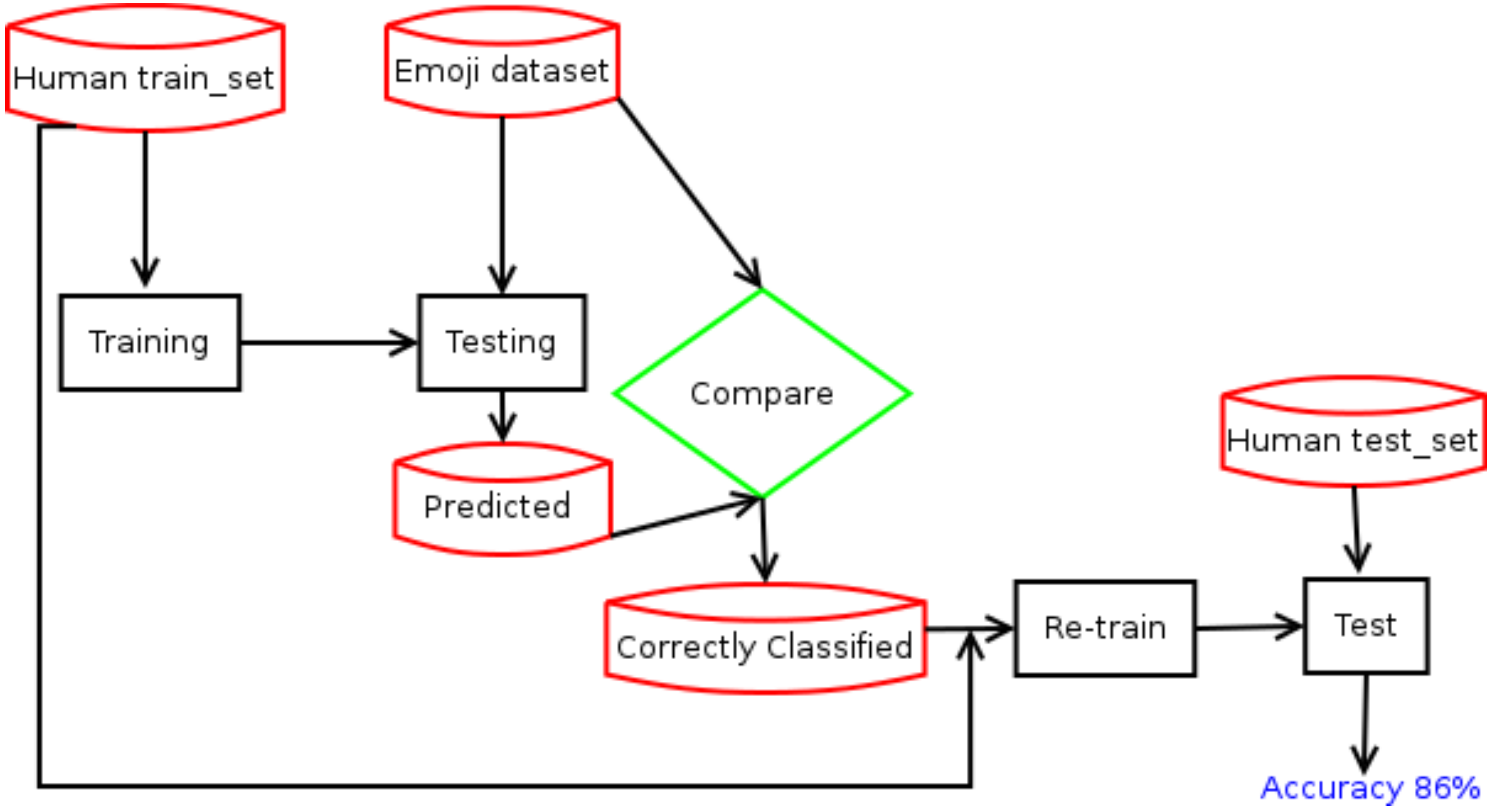
- This manual annotation process is time and money consuming.
- <u>**Soluation**</u>: Distant supervision or weak supervision . we use the emojis in the tweets to work as weak labels with which we can annotate the 36K tweets automatically.
  - Cons: not producing high quality dataset.
- Manually annotate 8k tweets as gold standard.
- To improve the quality of the automatic annotation and therefore the proposed tweets corpus
  - <u>**Soluation:**</u> <u>Self-training</u> approach is employed on the data to improve the classification and increase the accuracy of the annotation by exploiting the Manual dataset.

3.3. SELF TRAINING ON DISTANT SUPERVISION CORPUS

	Human annotated	Emojis annotated
	Label Distribution	
#Positive	3,705	14,468
#Negative	3,911	14,784
	Train/Test Distribution	
#Train_set	6,092	23,401
#Test_set	1,524	5,851
#Total_set	7,616	29,252

Statistics of the human annotation subset and the emojis distant supervision subset after subtract the human dataset

3.3. SELF TRAINING ON DISTANT SUPERVISION CORPUS



3.3. SELF TRAINING ON DISTANT SUPERVISION CORPUS

Experiment	#Train	#test	Baseline	Complex
Manual	6,092	1,524	71%	79%
Mixed	6,092	29,252	63%	76%
double-check	28,634	1,524	**77%**	**86%**
Non-check	35,341	1,524	70%	81%

The performance of the baseline and complex models on different datasets.

4. REPRODUCE RESULTS

- Pick up one paper and reproduce the result
  1. Deep learning Approach for Arabic SA (40 tweets) (slide 28)
  2. LSTM with an average accuracy of 81.31%
- Apply the same methods into different corpora

4. REPRODUCE RESULTS

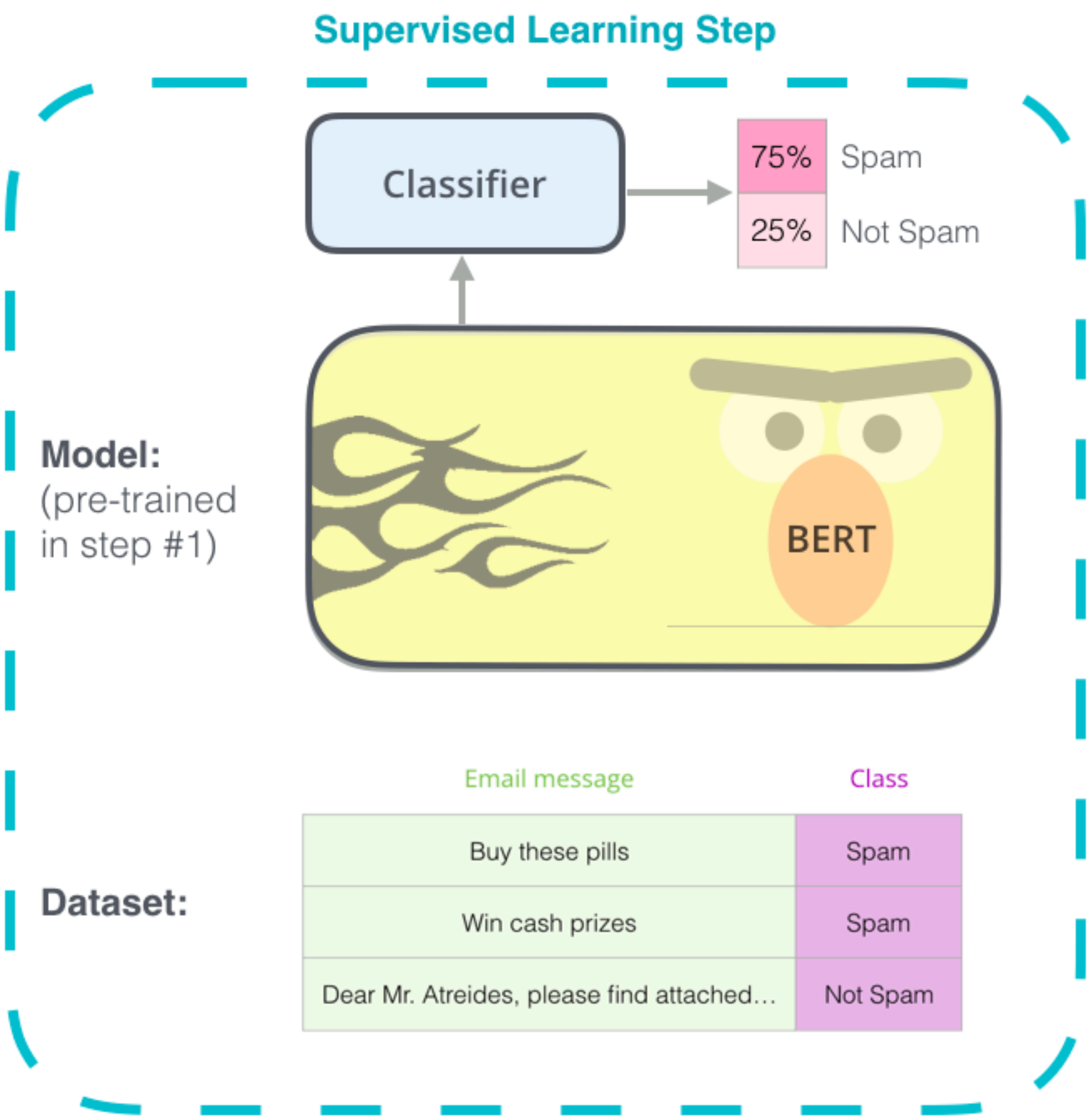
Corpus	Accuracy	MCC
40k Tweets	59%	0.19
ASTD	65%	0.09
ATSAD (our corpus)	53%	0

5. BERT

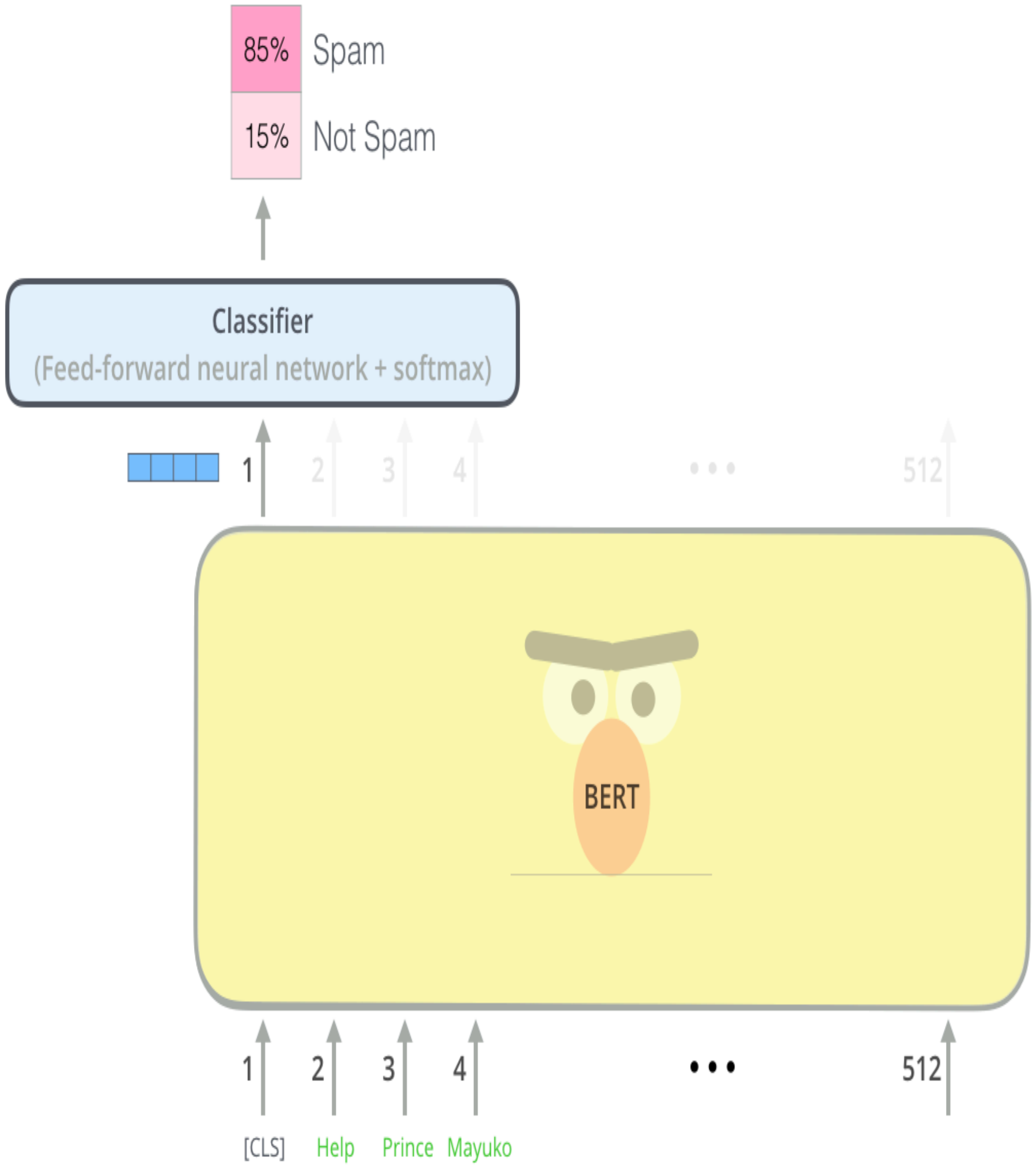
1. BERT (Bidirectional Encoder Representations from Transformers) is a recent paper published by researchers at Google AI Language.
2. As opposed to directional models, which read the text input sequentially (left-to-right or right-to-left), the Transformer encoder reads the entire sequence of words at once. Therefore it is considered bidirectional. This characteristic allows the model to learn the context of a word based on all of its surroundings (left and right of the word).

5. BERT FINE TUNNING

- BERT can be used for a wide variety of language tasks, while only adding a small layer to the core model:
- Classification tasks such as sentiment analysis are done by adding a classification layer on top of the Transformer output for the [CLS] token.



Email message	Class
Buy these pills	Spam
Win cash prizes	Spam
Dear Mr. Atreides, please find attached…	Not Spam



5. BERT FINE TUNNING

Epoch	10
Batch size	32
Max-Len	80
Data Split	60,20,20

5. BERT FINE TUNNING

Corpus	Accuracy	MCC
40k Tweets	83%	0.66
ASTD	82%	0.58
ATSAD (our corpus)	79%	0.58

REFERENCES

1. Al-Ayyoub, Mahmoud, et al. "Deep learning for Arabic NLP: A survey." *Journal of computational science* 26 (2018): 522-531.
2. R. Al-Jawfi, Handwriting Arabic character recognition LeNet using neural network, Int. Arab J. Inf. Technol. 6 (3) (2009) 304–309.
3. V. Jindal, A deep learning approach for Arabic caption generation using roots-words, AAAI (2017) 4941–4942.
4. Y.A. Alotaibi, Spoken Arabic digits recognizer using recurrent neural networks, in: Proceedings of the Fourth IEEE International Symposium on Signal Processing and Information Technology, 2004, IEEE, 2004, pp. 195–199.
5. A.E.-D. Mousa, H.-K.J. Kuo, L. Mangu, H. Soltau, Morpheme-based feature-rich language models using deep neural networks for LVCSR of Egyptian Arabic, in: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2013, pp. 8435–8439.
6. Y. Kim, Y. Jernite, D. Sontag, A.M. Rush, Character-aware neural language models, AAAI (2016) 2741–2749.
7. T. Deselaers, S. Hasan, O. Bender, H. Ney, A deep learning approach to machine transliteration, Proceedings of the Fourth Workshop on Statistical Machine Translation, Association for Computational Linguistics (2009) 233–241.
8. Y. Belinkov, J. Glass, A Character-Level Convolutional Neural Network for Distinguishing Similar Languages and Dialects, 2016 arXiv:1609.07568.
9. S. Malmasi, M. Zampieri, N. Ljubesˇic, P. Nakov, A. Ali, J. Tiedemann, Discriminating between similar languages and Arabic dialect identification: a report on the third DSL shared task, VarDial, vol. 3 (2016) 1.
10. Y. Samih, M. Attia, M. Eldesouki, A. Abdelali, H. Mubarak, L. Kallmeyer, K. Darwish, A neural architecture for dialectal Arabic segmentation, Proceedings of the Third Arabic Natural Language Processing Workshop (2017) 46–54.
11. Y. Yao, Z. Huang, Bi-directional LSTM recurrent neural network for Chinese word segmentation, in: International Conference on Neural Information Processing, Springer, 2016, pp. 345–353.
12. Abdullah, Malak, Mirsad Hadzikadicy, and Samira Shaikhz. "SEDAT: sentiment and emotion detection in Arabic text using CNN-LSTM deep learning." *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2018.
13. Heikal, Maha, Marwan Torki, and Nagwa El-Makky. "Sentiment analysis of Arabic tweets using deep learning." *Procedia Computer Science* 142 (2018): 114-122.
14. Mohammed, A., Kora, R. Deep learning approaches for Arabic sentiment analysis. *Soc. Netw. Anal. Min.* **9**, 52 (2019). https://doi.org/10.1007/s13278-019-0596-4
15. Abo, Mohamed Elhag Mohamed, Ram Gopal Raj, and Atika Qazi. "A Review on Arabic Sentiment Analysis: State-of-the-Art, Taxonomy and Open Research Challenges." *IEEE Access 7* (2019): 162008-162024.
16. Alwakid, Ghadah, Taha Osman, and Thomas Hughes-Roberts. "Challenges in sentiment analysis for Arabic social networks." *Procedia Computer Science* 117 (2017): 89-100.
17. Boudad, Naaima, et al. "Sentiment analysis in Arabic: A review of the literature." *Ain Shams Engineering Journal* 9.4 (2018): 2479-2490.
18. Alharbi, Amal, Mounira Taileb, and Manal Kalkatawi. "Deep learning in Arabic sentiment analysis: An overview." *Journal of Information Science* (2019): 0165551519865488.
19. https://www.blog.google/products/search/search-language-understanding-bert/
20. Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." *arXiv preprint arXiv:1810.04805* (2018).
21. https://towardsdatascience.com/bert-explained-state-of-the-art-language-model-for-nlp-f8b21a9b6270