International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075, Volume-8 Issue-8, June, 2019

# An Empirical Evaluation of Temporal Convolutional Network for Offensive Text Classification

**Murali S, Swapna T R**

***Abstract*: Concomitant with the ubiquity of digitization, the wide-open gate, for unfettered use of offensive content in digital platforms, is quite apparent. The veil of anonymity, offered by the digital platforms, has been misused by miscreants who engage in inappropriate, discourteous and rude conversations. The platform, which was meant for people to share their views, ideas constructively and collaborate for the collective betterment of society, is bombarded with offensive tweets with ulterior motives. There have been multiple efforts by various stakeholders to identify and classify such posts automatically employing different algorithms but the changing texting styles by different users have made this a challenging task. The Government of India (GoI) on August 2018 goaded the social media barons to voluntarily enforce sufficient guidelines for the content on their platforms, as the extant measures were rudimentary and inept.**

***While it has been established that the LSTM (Long Short-Term Memory) and the GRU (Gated Recurrent Unit) have been the state-of-the-art sequence modeling neural networks, the efficacy of Temporal Convolutional Network (TCN), which has been proposed as a viable alternate to LSTM/GRU for sequence modeling has not yet been explored for text (offensive) classification. In our work, we have evaluated the performance of TCN to identify and classify offensive language based on the intensity of its offensive content along with the conventional Convolutional Neural Network (CNN) and the state-of-the-art sequence modeling neural networks LSTM and GRU. Unlike LSTM and GRU, TCN exploits parallelism and is able to retain long range history with dilated convolutions and residual blocks. In addition, the TCN classifier was assessed for hate speech, aggression and harassment datasets. In all three datasets, the TCN set new benchmark scores (weighted F1).***

***Index Terms*: Deep Learning, Temporal Convolutional Network, Text Classification, Toxic.**

## I. INTRODUCTION

Nowadays, there is a plethora of platforms to informally share, discuss and collaborate in the ubiquitous digital social media. News portals, blogs, social media channels such as Twitter, Facebook, and YouTube enable media enthusiasts to express their views through conversational text messages, pictorial or graphical contents in the social media. Free and unhindered online expression of views or opinions, on any topic, has been significantly hampered these days with the advent of repugnant miscreants by their extremist, hateful, discriminative, abusive and harassing outbursts on digital platforms. Such threats inhibit people's unsullied exchange of fruitful conversations. Hence, automated identification and classification of displeasing, provocative online content has become a demanding and strenuous task in improving the common man's overall user experience. Such automatic classification will aid the digital platform and rejuvenate the quality usage of the platform as a tool for voicing out individuals' opinion.

The identification and categorization of offensive content has been heavily explored with classical Natural Language Processing (NLP) and Machine Learning techniques. But considering the constraints associated with the natural language such as word spell variations and typos, polysemy, contextual ambiguity and semantic variations, supervised machine learning technique turns out to be the most suitable technique for the task at hand. The enormous amount of data being generated from the digital platform, on a daily basis, further aids in favor of the supervised machine learning technique. Although offensive text classification naturally tends to be in the arena of traditional Natural Language Processing approach owing to the heavy reliability on the text, the traditional NLP approach fails to meet the enormous intricacies associated with the texting/commenting styles of different users, and has not proven to be scalable. Also, NLP approach is not a continuous learner, as new styles of adding comments (using symbols, mixture of both text and symbols) emerge, on a daily basis. Supervised Machine learning technique is a natural solution for a problem of such enormous depth and complexity. The latter has an inherent ability of continuous learning, with much shorter time compared to the classical NLP approach.

***We make the following original contributions in this work: a) empirical evaluation of Temporal Convolutional Network (TCN) for classification of text (for toxic comments) against the benchmark sequence models - LSTM and GRU. b) Set new benchmarks for Aggression, Hate Speech and Harassment classification by employing TCN.***

In subsequent sections, we discuss related work on classification of Toxic Comment/Hate Speech/Aggression, the utilized dataset, system description and conclude with discussion on results of our work.

## II. RELATED WORK

Multiple contributions have been made to identify and classify offensive textual content on digital platforms like Facebook and Twitter.

**Revised Manuscript Received on June 13, 2019.**

**Murali Sridharan**, Department of Computer Science and Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India.

**Swapna T R**, Department of Computer Science and Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India.

Spiros et al. [1] employed Convolutional Neural Network (CNN) with word2vec [2] word embedding and CNN without word embedding, for Kaggle's Toxic dataset. They compared the performance of both the CNN variants with the machine learning algorithms, which include Linear Discriminant Analysis (LDA), k Nearest Neighbor (kNN), Naïve Bayes (NB), and Support Vector Machine (SVM). The CNN model with text embedding proved to be their best performing model. Their results established the efficacy of CNN to capture the semantic features better than traditional machine learning algorithms. In the past, Chung et al. [3] empirically evaluated and concluded that the GRU is on par with LSTM for generic sequence modeling tasks (which include polyphonic music modeling and raw speech signal modeling). It must be noted that the authors did not make a concrete verdict if LSTM or GRU is superior based on their evaluation. Vishwanathan et al. [4] evaluated LSTM and GRU for comprehension modeling and established that the GRU outperforms LSTM in terms of higher classification accuracy for Question-Answering dataset. Bai, et al. [5] introduced the Temporal Convolutional Neural Network (TCN) and performed an empirical evaluation on a variety of generic tasks between Recurrent Neural Net(RNN) variants (LSTM, GRU, RNN) and TCN. They conducted an evaluation of generic tasks, which include synthetic stress tests, polyphonic music modeling, character-level language modeling, and word-level language modeling, with TCN as against the RNN, LSTM and GRU. TCN outperformed the RNN, LSTM and GRU in almost all the generic tasks. Kshirsagar, et al. [6] evaluated a Neural Network (NN) system for classification of Hate Speech. They employed a pre-trained word embedding, with a feed forward neural network, and achieved an F1 score of 0.924 on Hate Speech dataset [7] and achieved an F1 score of 0.932 on Harassment dataset [12]. Nowak, et al. [8] employed LSTM and GRU networks for classification of short text and sentiments. The LSTM network has the capability to remember long-range contextual history. Even though the vanishing gradient problem (as the gradient propagates back during the training, it becomes so small and finally vanishes before it could reach the early layers in the hierarchy, which would hamper the accuracy of the model and increase the training time significantly) is theoretically solved from vanilla RNN in LSTM, in practice LSTM is unable to efficiently handle long range history. The main reason behind this could be attributed to the multiple gates and sequential path between time steps; at each time step, the gradient slowly vanishes because of the product of the small gradient across each time step. In pursuit of a simpler mechanism with reduced gates, which could expediently process the data, the GRU was developed. In spite of GRU being much simpler in architecture complexity owing to the reduction in gates, it is established that it is only next to LSTM in terms of performance. It is vital to note that GRU is much more efficient and fast for smaller datasets or modelling sentences of smaller lengths. Kumar, et al. [9] created an annotated corpus for Aggression Identification task from Facebook comments and Twitter posts. Aroyehun, et al. [10] employed LSTM model and achieved a top weighted F1 score of 0.642 with data augmentation for the Aggression Identification of Facebook test set in TRAC -1 (Workshop on Trolling, Aggression and Cyber-bullying) at COLING 2018. Raiyani, et al [11] employed Dense Neural Networks (DNN) (Feed Forward Neural Network) with one hot representation achieved top score 0.60 for Aggression identification of Twitter test set in Trolling, Aggression and Cyber-bullying(TRAC).

## III. SYSTEM DESCRIPTION

The important features of TCN are: 1) causal convolution, that is, at time 't' the filter has convolved only till 't' but not later than 't'; 2) TCN can take any arbitrary length sequence

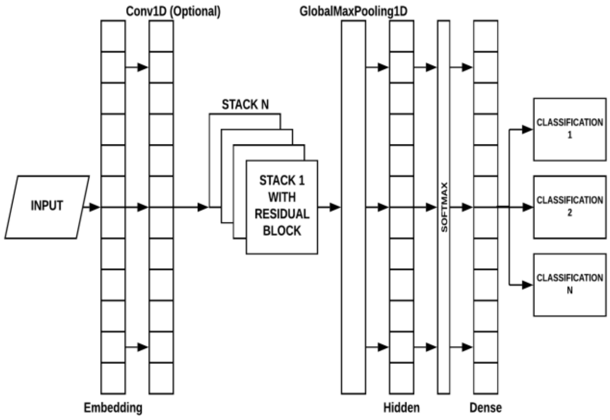


Fig. 1. Temporal Convolutional Network [3] for Text Classification

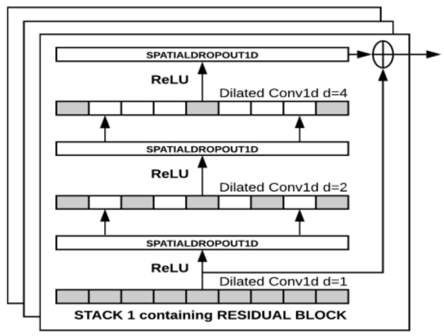


Fig. 2. Stack with Dilated Convoltion and Residual Block [3]

and turn it to an output sequence of the same arbitrary input length, similar to that of RNN. Beyond this the emphasis is to pursue long history context (which will focus on the capability of the network to contextualize far into the past to make prediction) utilizing a blend of deep neural nets (supplemented with residual layers) and dilated causal convolutions. As illustrated in Figure 1, the embedding layer is followed by an optional and the initial Convolution1D. Further, there are 'N' number of stacks, where each stack contains a residual block with dilated convolution filter, with dilation factor 'd' and SpatialDropout1D. The dilation factor could be varied to accommodate multiple sentences of varying length. Once several features have been extracted from various filter maps, the feature matrix is put forward to GlobalMaxPooling1D to derive the sub-sample of the most important features; subsequently, the sub-sampled data is fed to the final layer, which is the


*Published By:*
*Blue Eyes Intelligence Engineering*
*& Sciences Publication*

Fully Connected/Dense layer. Oord, et al. [13] employed dilated convolutions in order to expand the receptive field exponentially. The dilation factor controls the increase of receptive field (portion of the input sequence where the network is focusing to extract feature), which is crucial in ascertaining the breadth of the dilation. The important potential in a dilated convolution is that, instead of the sequential stride the filter skips and slides over the complete input sequence. Such a skip is determined by the dilation factor. Each dilation factor would stack up convolutional layer, which was formed by skipping and sliding over, corresponding to the respective dilation factor. This dilation factor is not constant and could be increased (in powers of two). By manipulating the dilation factor and stacking up the convolution layer (as illustrated in Figure 2), long range history could be effectively tapped, however long the sequence is, without the worry of vanishing gradient problem. This provides two ways to increase the receptive field of the TCN: choosing larger filter sizes k and expanding dilation factor d, where the effective history of one such layer is (k−1) d. A residual block contains two dilated convolutions, and the output, from the last convolution, is added with the direct input, to obtain the final output. This effectively allows layers to learn modifications to the identity mapping rather than the entire transformation, which has repeatedly been shown to benefit very deep networks. For a network with several layers, it is vital to ensure that the network does not overfit, hence the regularization of deep TCN is paramount to the efficiency of the neural net owing to increasing depth of the network N, dilation factor d and the size of the filter k, all three influence network's receptive field.

### A. Dataset

This experiment involved three datasets; the datasets were subject to the same data pre-processing steps (which include converting to lower case, then removing multiple space, new line characters, digits, URLs (Uniform Resource Locator), punctuation and other symbols) except for the Facebook corpus of Aggression dataset [9]. Pre-processing involved identification of comments in Hindi, followed by their translation into English, as the dataset had instances of both Hindi and English. Apart from translations, spell check and corrections were also implemented. The Toxic dataset from Kaggle's Toxic Comment Classification Challenge, 2018 is the choice of dataset (as it cuts across multiple facets of offensive categories) for the performance evaluation of TCN against variants of RNN, for modeling sequence data. The Toxic dataset contains a total of 1 59 570 instances of Wikipedia comments. They were annotated manually, for toxicity that included Identity Hate, Toxic, Obscene, Insult and Threat. The Hate Speech dataset [7] contains a total of 24 783 twitter posts. It contains a total of 1 430 hate speech instances and 23 353 instances of non-hate speech tweets. The Aggression dataset [9] has a total of 14 998 instances from the Facebook corpus. This dataset has a total of 6 284 clean instances, 5 296 of covertly aggressive, and 3 418 overtly aggressive instances. The test set has a total of 915 instances, obtained from the Facebook corpus, and 1 296 instances from the Twitter corpus. The harassment dataset [12] has a total of 20 360 instances, of which 5 285 belong to harassment category and 15 075 belong to non-harassment category.

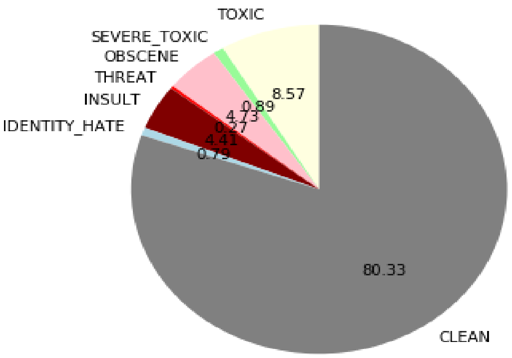


Fig. 3. Toxicity split (in %) from Toxic Dataset

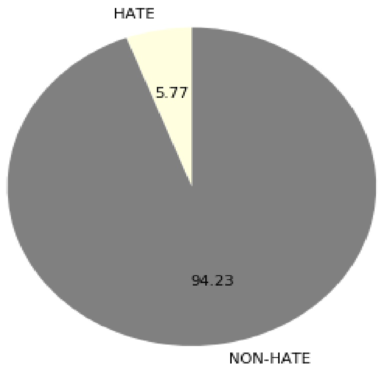


Fig. 4. Hate Speech split (in %) from Hate-Speech Dataset [5]

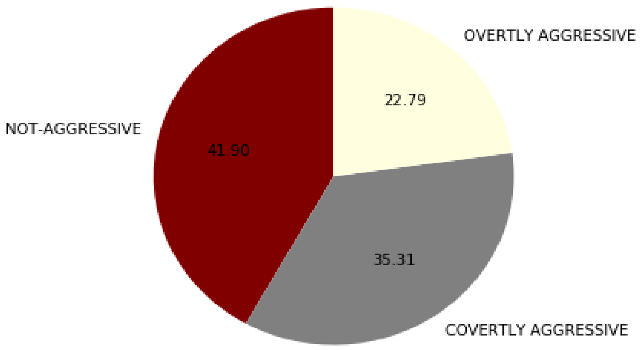


Fig. 5. Aggression split (in %) from Aggression dataset [7]

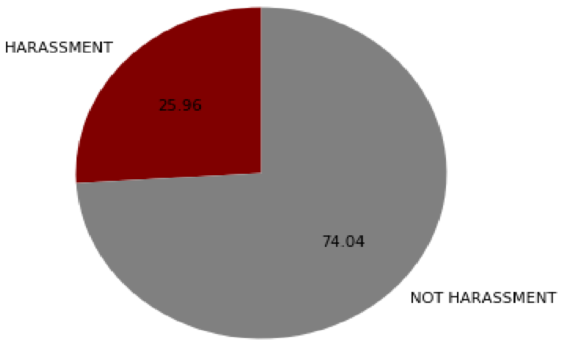


Fig. 6. Harassment split (in %) from Harassment dataset [10]


Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

## B. Environment

The generation, optimization and the evaluation of the entire Neural Network (NN) were carried out in an execution environment, which was equipped with NVIDIA GTX 1060 with CUDA compute capability of 6.1 and 16 GB RAM.

# IV. RESULTS

The datasets which were involved in this work have class imbalance, hence we have evaluated the Toxic dataset with both Weighted and Macro F1 scores. The other datasets were evaluated against their respective benchmark metric - Weighted F1. The main difference between the weighted and macro F1 is that the share of each label in the dataset is taken into account in the former whereas it is not taken into account in the latter.

## A. Empirical Evaluation of Toxic Dataset

Performance evaluation of the sequence data benchmark models such as LSTM and GRU (RNN variants) with the TCN and conventional CNN were carried out. It is important to note that the prime focus was to evaluate the performance of TCN for Toxic Comment Classification, as against the conventional CNN, and benchmark models for sequence data classification task. The binary cross entropy was used as the loss function and Adam [14] as the optimization function. Training loss is the classification (or prediction) error incurred during evaluation of the training dataset. Validation loss is the classification (or prediction) error incurred during appraisal of the validation dataset, through the trained network. Fig. 7 shows that the classification accuracy of TCN is almost matched by LSTM, but TCN achieved the 98.39% accuracy in much lesser time than the LSTM. Even though the GRU and CNN models did not show promising results in toxic comment classification, their classification accuracy lags behind the TCN and LSTM only by a miniscule percentage.

TABLE I. TOXIC COMMENT CLASSIFICATION MACRO AVERAGED F1

| MODEL | *CROSS-VALIDATION MACRO F1 in %* |
|---|---|
| CNN | 53.13 |
| LSTM | 59.61 |
| GRU | 56.87 |
| **TCN** | **62.66** |

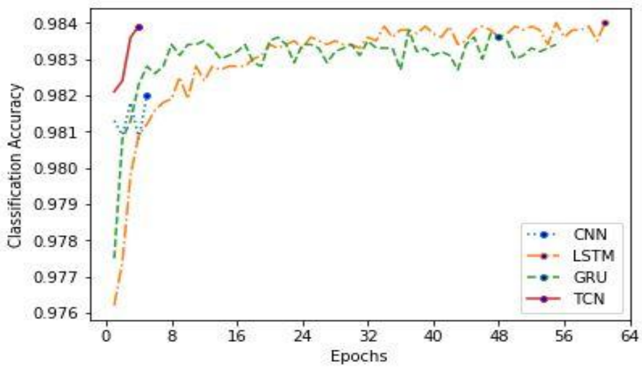


Fig. 7. Classification Accuracy Analysis – Toxic Comment Classification

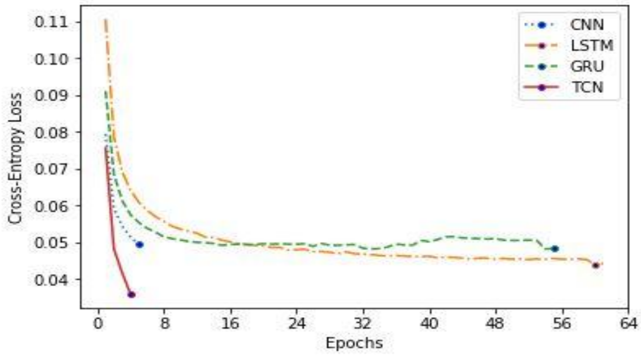


Fig. 8. Training Loss Analysis – Toxic Comment Classification

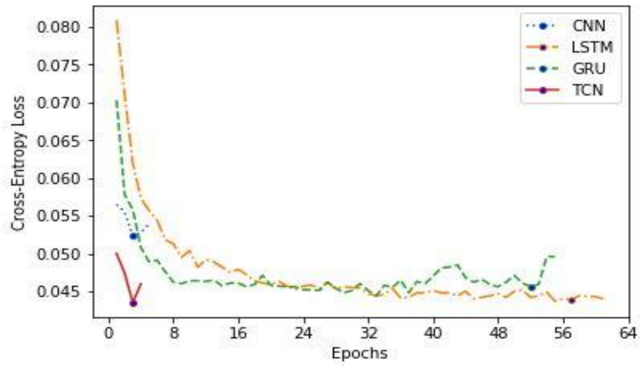


Fig. 9. Validation Loss Analysis – Toxic Comment Classification

From Fig. 8 and Fig. 9, it is evident that the CNN and TCN utilized their massive parallel computation ability and converged much quicker (in a few epochs) compared to the RNN variants: LSTM and GRU. TCN converged much quicker than CNN, with minimal loss compared to all the other NN models, indicative of its better generalization (or approximation) ability, compared to the other three models. CNN, LSTM and GRU converged, after multiple epochs, almost at the same loss value. Even though the RNN variants LSTM and GRU converged slower, both these models improvised on the loss function, in a steady manner. Similar to the training loss analysis, TCN converged much quicker during validation analysis as well. TCN and LSTM converged almost with same loss value, confirming that the LSTM model is effective for sequence modeling, whereas the TCN went ahead the generalization function of LSTM within a very few epochs, that is, in a very short time, an advantage which should be attributed to the inherent parallel processing capability of the CNN option. The sequence benchmark model variants LSTM and GRU gained in slowly after each epoch, that is, there was steady improvement of accuracy in successive epochs.

## B. Evaluation of Hate Speech Dataset

Table II depicts result for the hate speech classification. The TCN model achieved state of the art F1 score for the Hate Speech dataset, outperforming GRU [15] and Multi-Layer Perceptron (MLP)/Feed-Forward Neural Network [6] with embedding.


Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

TABLE II. HATE SPEECH CLASSIFICATION -F1 RESULTS

| MODEL | WEIGHTED F1 |
|---|---|
| GRU [15] | 0.890 |
| MLP + Word Embedding [6] | 0.924 |
| **TCN** | **0.933** |

TABLE III. HATE SPEECH CLASSIFICATION - CONFUSION MATRIX

| | NON-HATE | HATE |
|---|---|---|
| **NON-HATE** | 2296 | 19 |
| **HATE** | 137 | 27 |

Table III shows the confusion matrix for the Hate Speech classification. The high true positive score (2 296) could be credited to the higher number of non-hate speech instances in the training set.

### C. Evaluation of Aggression Dataset

Table IV contains the classification results for the Aggression section of the Facebook test set. The TCN model was trained with Facebook corpus. The trained model outperformed the top score for the Facebook test set (as indicated in Table IV) and trailed behind the top performing system for Twitter test set (as indicated in Table VI) by a miniscule amount.

TABLE IV. AGGRESSION CLASSIFICATION (FACEBOOK)–WEIGHTED F1 RESULTS

| MODEL | WEIGHTED F1 |
|---|---|
| LSTM [10] | 0.64 |
| **TCN** | **0.69** |

TABLE V. AGGRESSION CLASSIFICATION – CONFUSION MATRIX (FACEBOOK)

| | NOT AGGR | COVERTLY AGGR | OVERTLY AGGR |
|---|---|---|---|
| **NOT AGGR** | 460 | 136 | 33 |
| **COVERTLY AGGR** | 73 | 62 | 7 |
| **OVERTLY AGGR** | 40 | 51 | 53 |

Note: 'AGGR' is short form of 'AGGRESSIVE'

Table V shows the confusion matrix for the aggression classification within the Facebook corpus. It is evident that Not Aggressive category (NOT AGGR) is classified correctly followed by the Covertly Aggressive (Covertly Aggressive) and Overtly Aggressive (Overtly Aggressive) categories.

TABLE VI. AGGRESSION CLASSIFICATION (TWITTER) – WEIGHTED F1 RESULTS

| MODEL | WEIGHTED F1 |
|---|---|
| **MLP + Embedding [11]** | **0.60** |
| TCN | 0.59 |

Table VI highlights the Aggression classification results of the Twitter test set of the Aggression dataset [9]. Even though, the TCN model was trained with Facebook corpus, without any additional training data, the classification accuracy was trailing the top performing system by a very small difference.

TABLE VII. AGGRESSION CLASSIFICATION – CONFUSION MATRIX (TWITTER)

| | NOT AGGR | COVERTLY AGGR | OVERTLY AGGR |
|---|---|---|---|
| **NOT AGGR** | 381 | 56 | 45 |
| **COVERTLY AGGR** | 245 | 69 | 99 |
| **OVERTLY AGGR** | 119 | 81 | 161 |

Table VII contains the confusion matrix for classification of the Twitter test set. It can be observed that the Not Aggressive category is classified more correctly compared to the Covertly Aggressive category and Overtly Aggressive category respectively.

### D. Evaluation of Harassment Dataset

The TWEM (Transformed Word Embedding Model) [6] achieved a maximum 0.71 F1 score for Harassment dataset [12].

TABLE VIII. HARASSMENT CLASSIFICATION – WEIGHTED F1 RESULTS

| MODEL | WEIGHTED F1 |
|---|---|
| MLP + Word Embedding [6] | 0.71 |
| **TCN** | **0.73** |

TCN outperforms the established benchmark with 0.73 as listed in Table VIII.

TABLE IX. HARASSMENT CLASSIFICATION - CONFUSION MATRIX

| | NON-HARASSMENT | HARASSMENT |
|---|---|---|
| **NON-HARASSMENT** | 1396 | 110 |
| **HARASSMENT** | 384 | 146 |

Even though the number of Harassment instances in the dataset is comparatively low to that of Non Harassment instances, TCN has performed better with low false negatives (110) and false positives (384) as observed in Table IX.

## V. CONCLUSION

Conventionally, CNNs were considered to extract features from video frame or regular images, while Recurrent Neural Nets (RNNs) excelled at sequence problems, which include text and speech problems, predominantly. Major enhancements to vanilla RNN, including GRU and LSTM, boosted the long


*Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication*

range history contextualization but they did not turn out to be optimal, and effective in real time. With TCN, long range history could be exploited much better, in real world scenarios, and the word-word correlation could also be captured effectively. From the evaluation results, noted herein, it is evident that Temporal Convolution Neural Network outperforms conventional CNN LSTM and GRU models for offensive text classification. Furthermore, TCN performed better than RNN variants LSTM and GRU, with respect to the macro averaged F1 scores of Toxic Comment Classification, in much lesser training time than the RNN variants: LSTM and GRU. GRU, considered to be the most effective model for use in short and medium sentences (<500 words), is trailing behind TCN and LSTM in Toxic Comment Classification task. We conclude that, TCN equipped with Residual Blocks, Causal and Dilated Convolution is able to model long range sentence and is able to capture the long range context much more efficiently, compared to all the other models, as observed in this study. The effectiveness of TCN for offensive text classification is further demonstrated by extended classification tests of Hate Speech, Aggression and Harassment datasets. Of the three extended offensive text classification under observation, TCN set new benchmarks and performed better than the established sequence models - LSTM and GRU. Further to this evaluation, we intend to experiment TCN with Embeddings from Language Models (ELMo) equipped with deep bi-directional language model, which is pre-trained on a very large corpus; and analyze the classification effectiveness for multiple datasets.

## ACKNOWLEDGMENT

We acknowledge the contribution and guidance of Prof. Latha Parameshwaran, Computer Science and Engineering Department, Amrita Vishwa Vidyapeetham.

## REFERENCES


1. Georgakopoulos, Spiros V., Sotiris K. Tasoulis, Aristidis G. Vrahatis, and Vassilis P. Plagianakos. "Convolutional neural networks for toxic comment classification." In Proceedings of the 10th Hellenic Conference on Artificial Intelligence, p. 35. ACM, 2018.
2. Goldberg, Yoav, and Omer Levy. "word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method." arXiv preprint arXiv:1402.3722 (2014).
3. Chung, Junyoung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. "Empirical evaluation of gated recurrent neural networks on sequence modeling." arXiv preprint arXiv:1412.3555 (2014).
4. Viswanathan, Sujith, M. Anand Kumar, and K. P. Soman. "A Sequence-Based Machine Comprehension Modeling Using LSTM and GRU." In Emerging Research in Electronics, Computer Science and Technology, pp. 47-55. Springer, Singapore, 2019.
5. Bai, Shaojie, J. Zico Kolter, and Vladlen Koltun. "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling". arXiv preprint arXiv:1803.01271 (2018).
6. Kshirsagar, Rohan, Tyus Cukuvac, Kathleen McKeown, and Susan McGregor. "Predictive Embeddings for Hate Speech Detection on Twitter." arXiv preprint arXiv:1809.10644 (2018).
7. Davidson, Thomas, Dana Warmsley, Michael Macy, and Ingmar Weber. "Automated hate speech detection and the problem of offensive language." In Eleventh International AAAI Conference on Web and Social Media. 2017.
8. Nowak, Jakub, Ahmet Taspinar, and Rafał Scherer. "LSTM recurrent neural networks for short text and sentiment classification." In International Conference on Artificial Intelligence and Soft Computing, pp. 553-562. Springer, Cham, 2017.
9. Kumar, Ritesh, Aishwarya N. Reganti, Akshit Bhatia, and Tushar Maheshwari. "Aggression-annotated corpus of hindi-english code-mixed data." arXiv preprint arXiv:1803.09402 (2018).
10. Aroyehun, Segun Taofeek, and Alexander Gelbukh. "Aggression detection in social media: Using deep neural networks, data augmentation, and pseudo labeling." In Proceedings of the First Workshop on Trolling, Aggression and Cyberbullying (TRAC-2018), pp. 90-97. 2018.
11. Raiyani, Kashyap, Teresa Gonçalves, Paulo Quaresma, and Vítor Nogueira. "Fully connected neural network with advance preprocessor to identify aggression over facebook and twitter." (2018).
12. Golbeck, J., Ashktorab, Z., Banjo, R.O., Berlinger, A., Bhagwan, S., Buntain, C., Cheakalos, P., Geller, A.A., Gergory, Q., Gnanasekaran, R.K. and Gunasekaran, R.R., 2017, June. A large labeled corpus for online harassment research. In Proceedings of the 2017 ACM on Web Science Conference (pp. 229-233). ACM.
13. Oord, Aaron van den, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. "Wavenet: A generative model for raw audio." arXiv preprint arXiv:1609.03499 (2016).
14. Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." arXiv preprint arXiv:1412.6980(2014).
15. Founta, Antigoni-Maria, Despoina Chatzakou, Nicolas Kourtellis, Jeremy Blackburn, Athena Vakali, and Ilias Leontiadis. "A unified deep learning architecture for abuse detection." arXiv preprint arXiv:1802.00385 (2018).


## AUTHORS PROFILE

**Murali S** is pursuing his final year MTech (CSE) from School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore. He completed his B.Tech from Anna University, Chennai in 2009. He has 8+ years of experience spanning Machine Learning/Deep Learning (1 year) for an automobile major; Application Development and Maintenance space (7 years) across multiple verticals and Quality (0.5 year) for an e-commerce giant. His primary focus for NLP & Computer Vision: includes neural network modelling, optimizing neural networks, generalization and transfer learning; for Application Development & Maintenance: includes requirements verification and feasibility, planning, design, implementation, delivery and support for the deliverables; for Quality: includes impact analysis, test plan creation, audit plan, audit, test case execution, documentation and report generation.

**Dr. Swapna T. R.** is an Assistant Professor in the Department of Computer Science & Engineering, School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore and has 17 years of teaching experience. She completed Ph. D. from the School of Medical Science and Technology (SMST), IIT Kharagpur, in the area of Computer Vision for Medical Imaging and MTech in Computer Science and Engineering from NIT-Calicut. Her research interests are in the area of Medical Imaging and Machine Learning.

Swapna has published in various Scopus-indexed international journals, conferences and edited book volumes including Springer and Elsevier. She teaches courses in Machine Learning, Modeling and Visualization, Computer Graphics, Medical Imaging, Artificial Intelligence, Data Mining, Computer Graphic and Big Data Analytics.