Troll Meme Classification for Tamil Memes Using Convolutional Neural Network

K. Nithya[1][0000-0002-0796-342X], **S. Sathyapriya**[1]**, M. Sulochana**[1]**, S. Thaarini**[1] **and C.R. Dhivyaa**[1]

[1] Department of Computer Science and Engineering, Kongu Engineering College, Tamilnadu, India

knithya89@gmail.com

Abstract. Memes have become a new type of internet communication. It has the ability to instantly disseminate anger, offensiveness, and violence. Because of its regional meaning, classifying memes is difficult.. This work presents here a computational model of classifying Tamil memes using convolutional neural networks. Convolutional neural networks have the potential to learn, adapt, and rearrange themselves. As a result, it can extract features automatically applying prior knowledge of existing categories, avoiding the time-consuming feature extraction process used in older methods in images. The basic layer of MobileNet is made up of depth-wise separable filters, also referred to as depth-wise separable convolution. The network structure is another feature that boosts performance. It utilizes very less than computation power while applying transfer learning. This network has reduced parameters and computation cost. Skip connections, or shortcuts, are used by residual neural networks to jump past some layers. Residual connections allow parameter gradients to travel more easily from the output layer to the network's prior layers, allowing for the training of deeper networks. Higher accuracies on more demanding tasks may arise from the greater network depth. AlexNet is a leading architecture for any image identification task, and it could have a lot of applications in the artificial intelligence field of computer vision. In the future, AlexNet may be used for image classification jobs more than CNNs. This work aims to classify Tamil memes using Mobilenet, Resnet and hyper parameter turned AlexNet.

Keywords: troll meme, convolutional neural network, mobilenet, resnet, alexnet.

1. Introduction

Memes are very much popular in social media due to their humorous means of conveying things to people. Memes are a type of culturally relevant expression. A meme is a picture or video that symbolizes a specific audience's thoughts and feelings. It is employed as a tool to draw people's attention to a current issue. The content of traditional media such as television, radio, and newspapers are monitored and analyzed. Social media platforms, on the other hand, allow internet users to communicate and contribute to their online communities without the need for moderation. Government policies and politics, as well as some myths circulated with the goal of mocking religious faith, have negative consequences on society. Although most of these internet users are harmless, due to the anonymity and freedom given by social media, some of them produce harmful information.

Propensity of to self-replicate and spread across cultures, memes have become an integral aspect of online communication. Most of these memes are humorous, but some of them may cross the line into being disrespectful to specific individuals or groups; these memes are known as troll memes. Troll memes are used by the opposition to embarrass, discredit, or distribute false information about famous people or government projects by disrespecting them harshly, which becomes a means of propagating one's animosity to the general public. Troll memes are never meant to be funny, but rather to ridicule and embarrass another person on the internet. Trolls are cruel and aim to make people laugh or mock them.

In this article 2nd Section describes some existing works implemented so far. 3rd Section explains the dataset which is chosen for our experiment. 4th Section provides details of the proposed system. 5th Section shows experimental results and 6th section gives conclusion and future work.

2. Literature survey

Amalia Amalia et al, in categorizing Indonesian government-based meme [1] used recognition of optical characters with Tesseract for the process of recognizing character through English alphabet dictionary [2] which transforms character in meme to be readable by machine which by Natural Language Processing (NLP) and Image Processing [3] to be applied on text separated images is fed to Naive Bayes which correctly categorizes up to 75%. Few tokens are not detected correctly by this approach.

Nikhil Kumar Ghanghor et al,2021 paper of Tamil, Malayalam, and Kannada Offensive Language Identification and Meme Classification using the dataset of (Suryawanshi et al.,2020b) for the Tamil language. Inception and ResNet50 were used for picture modality [4], and mBERT-uncased, mBERT-cased, XLMR, and IndicBERT were used for text modality. XLMR, IndicBERT, mBERT-uncased, mBERT-cased, Inception and Resnet algorithms are compared through accuracy. Finally, they conclude Multilingual BERT-cased was the best model for Tamil troll meme categorization, with a F1 score of 0.55. This multilingual BERT-cased model for classifying aggressive languages had weighted average F1 values of 0.75 in Tamil, 0.95 in Malayalam, and 0.71 in Kannada. Here, they achieved a 55% accuracy rate for mBERT-cased. Because memes are multimodal in nature, unimodal methods cannot be implemented for such cases.

Deep learning for meme classification author Manoj Balaji J, Chinmaya HS in the year of 2021 uses the dataset 'Tamil Troll Memes' Suryawanshi et al. (2020a) and algorithm for classifying memes is Resnet-50. The network can grasp features from different portions and levels of the image because of the architecture's 50 layers [5]. The photos are downsized to 64x64 pixels, and all three-color channels are input into the algorithm, which is then given to the neural network's subsequent layers to produce the final output. The resnet-50 has a 48% accuracy in classifying memes. The probability value for each class is determined via Softmax activation in the last layer, and the highest probability class is chosen as the output. The image's text/caption is in Tamil but written in Italic, and this paper's limitation is a mix of Tamil and English.

Jaeheon Kim et al, Understanding and identifying the use of emotes in toxic [6] chat on Twitch. NLP (natural language processing) algorithms [7] are used in the paper. Dataset taken from the Twitch platform. Accuracy of the paper is 74% and the drawback of the paper is Not including localized language.

Deep Convolutional Neural Networks for ImageNet Classification author Alex Krizhevsky et al, uses the ILSVRC 2010 and ILSVRC 2012 dataset. We used a regularization method named "dropout" to reduce overfitting in fully connected layers [8], which proved to be quite effective. This is a significant improvement above the prior state-of-the-art. The neural network is composed of five convolutional layers, two max pooling layers, three fully connected layers, and a final 1000-way softmax layer, which has parameters of about 60 million and 650,000 neurons [9]. And yield top1 to top5 error rates 37.5% and 17% at ILSVRC 2010 and 15.3% and 26.2% at ILSVRC 2012 dataset.

Wei Wang et al, 2020 in his Approach of Image Classification Using Dense model of MobileNet is an article that describes a novel approach to image classification using dense-mobilenet [10]. Dense

blocks proposed in DenseNet are added into MobileNet with the motive of reducing the number of network parameters and increasing the accuracy of classification[11].Same size input featured Convolutional network maps in MobileNet models with dense blocks in dense-MobileNet models, and perform the dense connections. They propose two Dense-MobileNet structures in this paper: Dense1-MobileNet and Dense2-MobileNet. The depthwise convolution layer and the point convolution layer are two independent convolution layers in this model[12]. Dense2 has two depthwise and two pointwise convolutional layers. Dense2-MobileNet can help improve classification accuracy. The classification accuracy of Dense1-MobileNet is less than that of MobileNet, the number of parameters is lower.Fairly Dense2-Mobilenet accuracy was 95% it is better than Mobilenet [13].

3. Overview of dataset

Dataset contains two thousand five hundred images of troll and non-troll tamil memes which are used as training dataset and the sample figures are shown in Figure 1 and Figure 2. The Dataset for troll tamil meme classification chosen from codalab. And for testing we have three hundred tamil memes and for training we have two thousand three hundred. Dataset containing troll memes are placed in a separate folder. Likewise, the dataset containing non troll memes are placed in a separate folder for training.

3.1. Data augmentation

In order to improve performance of the model and avoid subsequent overfitting the dataset needs to be augmented and its sample is shown in Figure 3. There are five popular data augmentation techniques used for image dataset augmentation. These include adjusting zoom, flip, random, rotation, random shift, brightness of the image for the dataset chosen. In addition to this we did rescaling and shearing of images [14][15].

Figure 1. Sample for non-troll meme.   **Figure 2.** Sample for troll meme.

Figure 3. Sample augmented image.

4. Proposed system

This work focused on proposing a better convolution neural network model for classifying tamil memes. Mobilenet a streamlined that uses depth wise separable convolution network to construct lightweight deep convolutional neural networks, Resnet a backbone for many computer vision tasks that allowed us to train deep neural networks with 150 + layers successfully and Alexnet first architecture of using convolutional layers in a sequential order The network's final fully connected layer

contains a softmax activation function that generates a vector that represents a probability distribution over 1000 classes and constructs a lightweight conventional neural network using depth wise separable convolutions to improve accuracy and yield efficient results [16][17].

- AlexNet performs better than Mobilenet and Resnet in image classification and alexnet recognizes characters in image and till now Alexnet is not used in Tamil meme classification.
- The AlexNet architecture is hyper turned to get slightly improved performance and accuracy.
- In order to improve accuracy, we have implemented training dataset augmentation.

5. Experiments

The AlexNet architecture is made up of several more convolutional layers, a max pooling layer, a normalisation layer, a conv-pool-norm layer, numerous fully linked layers, and then a few more convolutional layers which are shown in Figure 4. There are five conv layers and two fully connected layers prior to the final completely connected layer that is connected to the output classes. AlexNet was trained on ImageNet using 224x224x3 image inputs. This first layer, which is the AlexNet's conv layer, has 11x11 filters, with 96 of them applied. The output was 55x55x96, and the first layer had 35K parameters.
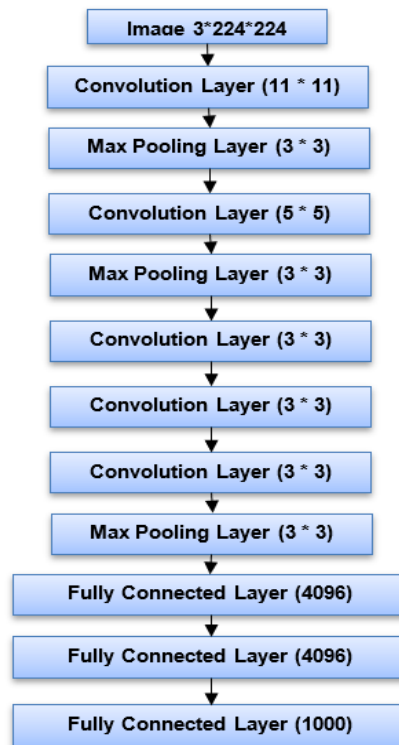
Figure 4. AlexNet architecture.

The second layer is a pooling layer, and at stride 2 we have three 1x1 filters applied. The max pooling layer's output volume is 27x27x96, with no parameters to learn. Because the parameters are the weights that are attempting to learn, they do not learn anything. We learn the weights of convolutional layers, but all we do with pooling is set a rule, look at the pooling region, and take the maximum [18][19]. As a result, there are no learnt parameters. At first, there are 11 by 11 filters, then five by five filters, and finally one by one filters. Finally, we have two fully connected layers. FC8 is the last layer, which connects to the Softmax and the 1000 ImageNet classes [20][21].
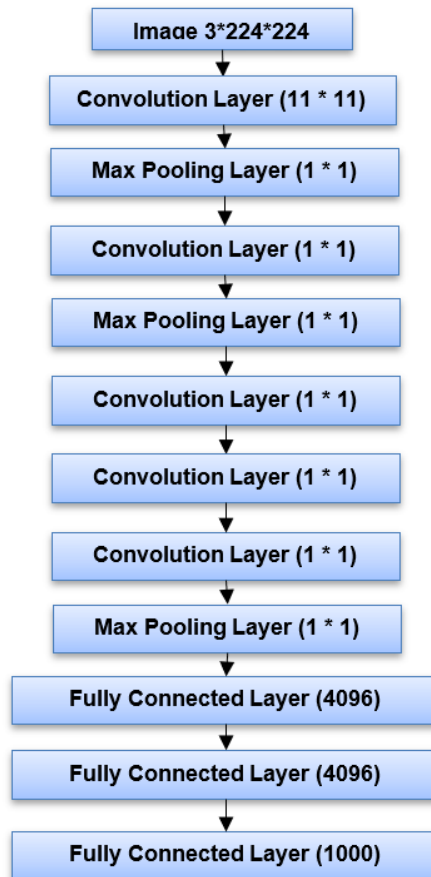
Figure 5. Modified alexnet architecture.

Fig 5 Modified AlexNet architecture is similar to that of AlexNet. It starts with a convolution layer, then moves on to a max pooling layer and then moves on batch normalization, conv - pool - norm, and finally a few convolution layers, a pooling layer, and numerous fully linked layers [22]. Before the last fully connected layer goes to the output class, there are five convolution layers and three completely connected layers. We used sigmoid activation instead of softmax activation to achieve better performance. ImageNet was used to train modified AlexNet, which had inputs of 227×227x3 images [23].

6. Result and conclusion
In order to predict accuracy after training the model with the dataset testing was done and accuracy was obtained.
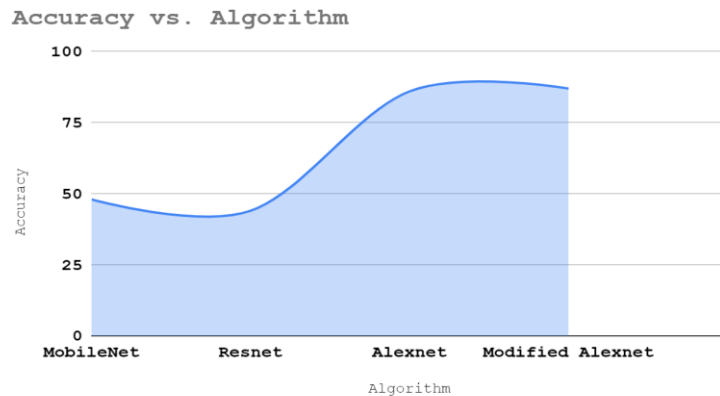
Figure 6. Accuracy of various CNN models for troll meme classification.
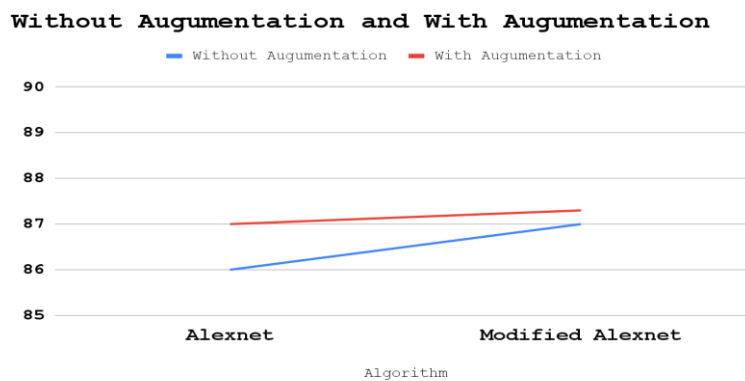

Figure 7. Accuracy increased with augmentation.

AlexNet performs better than MobileNet and Resnet. Modified AlexNet is a hyper parameter turned AlexNet which performs slightly better than AlexNet. Figure 6 shows troll meme classification using CNN models in terms of accuracy. Modified AlexNet shows better accuracy of about 87% then the rest of the algorithms.

Figure 7 Accuracy increased with Augmentation shows Troll meme classification using alexnet and modified alexnet in terms of accuracy. By using augmented dataset both alexnet and modified alexnet performs better than when trained with actual dataset. Thus, modified alexnet performs better in troll meme classification than various CNN models like MobileNet and ResNet.Even though convolutional neural network models like mobilenet, Resnet and Alexnet recognize characters even if they are in tamil as mentioned in the literature survey there are a few contradictions in predicting untrained memes with different text. For classifying troll memes with Modified AlexNet provides an accuracy of 87%.

Future Work. Future work involves using more hyper parameterized factors in Alex net, increasing the training dataset and training the model with the rest of the convolutional network model so that accuracy can be improved in predicting whether the meme as it is troll or not.

References

[1]   A. Amalia, A. Sharif, F. Haisar, D. Gunawan and B. B. Nasution, "Meme Opinion Categorization by Using Optical Character Recognition (OCR) and Naïve Bayes Algorithm," 2018 Third International Conference on Informatics and Computing (ICIC), 2018, pp. 1-5, doi: 10.1109/IAC.2018.8780410.

[2]   A. A. Marouf, R. Ajwad and A. F. Ashrafi, "Looking Behind the Mask: A framework for Detecting Character Assassination via Troll Comments on Social media using Psycholinguistic Tools," 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), 2019, pp. 1-5, doi: 10.1109/ECACE.2019.8679154.

[3]   H. Hitkul, R. R. Shah, P. Kumaraguru and S. Satoh, "Maybe Look Closer? Detecting Trolling Prone Images on Instagram," 2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM), 2019, pp. 448-456,doi: 10.1109/BigMM.2019.00029.

[4]   Ankit Kumar Mishra and Sunil Saumya. 2021. IIIT_DWD@EACL2021: Identifying Troll Meme in Tamil using a hybrid deep learning approach. In Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages, pages 243–248, Kyiv. Association for Computational Linguistics.

[5]   H. Tahir, M. Shahbaz Khan and M. Owais Tariq, "Performance Analysis and Comparison of Faster R-CNN, Mask R-CNN and ResNet50 for the Detection and Counting of Vehicles," 2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), 2021, pp. 587-594, doi: 10.1109/ICCCIS51004.2021.9397079.

[6]   Jan Kocoń, Alicja Figas, Marcin Gruza, Daria Puchalska, Tomasz Kajdanowicz, Przemysław Kazienko, Offensive, aggressive, and hate speech analysis: From data-centric to human- centered approach,Information Processing & Management, Volume 58, Issue 5, 2021, 102643, ISSN 0306-4573,https://doi.org/10.1016/j.ipm.2021.102643.

[7]   Jaeheon Kim, Donghee Yvette Wohn, Meeyoung Cha, Understanding and identifying the use of emotes in toxic chat on Twitch, Online Social Networks and Media,Volume 27, 2022, 100180, ISSN 2468-6964, https://doi.org/10.1016/j.osnem.2021.100180.

[8]   M. J. Dupuis and A. Williams, "The Spread of Disinformation on the Web: An Examination of Memes on Social Networking," 2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), 2019, pp. 1412-1418, doi: 10.1109/SmartWorld-UIC-ATC-SCALCOM-IOP-SCI.2019.00256.

[9]   L. Xiaolin, R. C. Panicker, B. Cardiff and D. John, "Multistage Pruning of CNN Based ECG Classifiers for Edge Devices," 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2021, pp. 1965-1968, doi: 10.1109/EMBC46164.2021.9630588.

[10]   H. Yanagisawa, T. Yamashita and H. Watanabe, "A study on object detection method from manga images using CNN," 2018 International Workshop on Advanced Image Technology (IWAIT), 2018, pp. 1-4, doi: 10.1109/IWAIT.2018.8369633.

[11]   Sathishkumar V E, Changsun Shin, Youngyun Cho, "Efficient energy consumption prediction model for a data analytic-enabled industry building in a smart city", Building Research & Information, Vol. 49. no. 1, pp. 127-143, 2021.

[12]   Sathishkumar V E, Youngyun Cho, "A rule-based model for Seoul Bike sharing demand prediction using Weather data", European Journal of Remote Sensing, Vol. 52, no. 1, pp. 166-183, 2020.

[13]   Sathishkumar V E, Jangwoo Park, Youngyun Cho, "Seoul Bike Trip duration prediction using data mining techniques", IET Intelligent Transport Systems, Vol. 14, no. 11, pp. 1465-1474, 2020.

[14]   Sathishkumar V E, Jangwoo Park, Youngyun Cho, "Using data mining techniques for bike sharing demand prediction in Metropolitan city", Computer Communications, Vol. 153, pp. 353-366, 2020.

[15]   Sathishkumar V E, Yongyun Cho, "Season wise bike sharing demand analysis using random forest algorithm", Computational Intelligence, pp. 1-26, 2020.

[16] Sathishkumar, V. E., Wesam Atef Hatamleh, Abeer Ali Alnuaim, Mohamed Abdelhady, B. Venkatesh, and S. Santhoshkumar. "Secure Dynamic Group Data Sharing in Semi-trusted Third Party Cloud Environment." Arabian Journal for Science and Engineering (2021): 1-9.

[17] Chen, J., Shi, W., Wang, X., Pandian, S., & Sathishkumar, V. E. (2021). Workforce optimisation for improving customer experience in urban transportation using heuristic mathematical model. International Journal of Shipping and Transport Logistics, 13(5), 538-553.

[18] Pavithra, E., Janakiramaiah, B., Narasimha Prasad, L. V., Deepa, D., Jayapandian, N., & Sathishkumar, V. E., Visiting Indian Hospitals Before, During and After Covid. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 30 (1), 111-123, 2022.

[19] Easwaramoorthy, S., Moorthy, U., Kumar, C. A., Bhushan, S. B., & Sadagopan, V. (2017, January). Content based image retrieval with enhanced privacy in cloud using apache spark. In International Conference on Data Science Analytics and Applications (pp. 114-128). Springer, Singapore.

[20] Sathishkumar, V. E., Agrawal, P., Park, J., & Cho, Y. (2020, April). Bike Sharing Demand Prediction Using Multiheaded Convolution Neural Networks. In Basic & Clinical Pharmacology & Toxicology (Vol. 126, pp. 264-265). 111 RIVER ST, HOBOKEN 07030-5774, NJ USA: WILEY.

[21] Subramanian, M., Shanmuga Vadivel, K., Hatamleh, W. A., Alnuaim, A. A., Abdelhady, M., & VE, S. (2021). The role of contemporary digital tools and technologies in Covid-19 crisis: An exploratory analysis. Expert systems.

[22] Sathishkumar, V. E., Rahman, A. B. M., Park, J., Shin, C., & Cho, Y. (2020, April). Using machine learning algorithms for fruit disease classification. In Basic & clinical pharmacology & toxicology (Vol. 126, pp. 253-253). 111 RIVER ST, HOBOKEN 07030-5774, NJ USA: WILEY.

[23] Sathishkumar, V. E., & Cho, Y. (2019, December). Cardiovascular disease analysis and risk assessment using correlation based intelligent system. In Basic & clinical pharmacology & toxicology (Vol. 125, pp. 61-61). 111 RIVER ST, HOBOKEN 07030-5774, NJ USA: WILEY.