EmoNet: Unveiling Affective States through Convolutional Neural Networks in Textual Emotion Classification

Said Salloum
School of Science, Engineering, and Environment
University of Salford
Manchester, UK
salloum78@live.com

Khalaf Tahat
Media & Creative Industries, United Arab Emirates University and Yarmouk University, Jordan
k.tahat@uaeu.ac.ae

Ahmed Mansoori
Media & Creative Industries department, United Arab Emirates University, UAE
a.mansoor@uaeu.ac.ae

Raghad Alfaisal
Faculty of Computing and Meta-Technology, Universiti Pendidikan Sultan Idris, Tanjung Malim, Malaysia
raghad.alfaisal81@gmail.com

Dina Tahat
Al Ain University, Applied sociology department, UAE
dina.tahat@aau.ac.ae

***Abstract*—In the burgeoning field of natural language processing, emotion classification from textual data has emerged as a critical task with applications ranging from sentiment analysis to mental health assessment. This paper explores the utilization of Convolutional Neural Networks (CNNs), traditionally dominant in image processing, for classifying emotions in text. Our proposed CNN model leverages the inherent hierarchical structure of language to identify and learn emotion-specific features, with an emphasis on capturing contextual n-grams through convolutional filters. The approach is substantiated by a comprehensive dataset, subjected to rigorous preprocessing and vectorization via TF-IDF to convert text into a numerical format suitable for deep learning. The model's architecture is meticulously crafted, incorporating convolutional layers followed by global max pooling and dense layers, culminating in a softmax activation function tailored for multi-class classification. Our findings demonstrate the model's robustness, achieving a notable accuracy of 96.08% on the test set. This high level of precision is further corroborated by the Receiver Operating Characteristic (ROC) analysis, revealing exceptional area under the curve (AUC) values across various emotion categories. The results suggest that CNNs hold significant promise for emotion recognition tasks in textual data, providing an effective framework for future explorations in the domain.**

Keywords—Convolutional Neural Networks (CNNs), Emotion Classification, Natural Language Processing (NLP), Receiver Operating Characteristic (ROC).

I. Introduction

Emotion classification in text is a pivotal task in the field of natural language processing (NLP) that has garnered considerable attention in recent years [1], [2] . The ability to accurately discern emotions from text has profound implications across a myriad of applications, from enhancing user experience in digital assistants to monitoring mental health through social media analysis [3], [4]. With the exponential increase in digital textual content, the need for automated emotion recognition systems has become ever more pressing [5]–[7].

Deep learning has revolutionized the landscape of NLP, offering powerful tools that excel in capturing complex patterns in data [8]. Convolutional Neural Networks (CNNs), in particular, have shown exceptional performance in various NLP tasks, traditionally being the go-to model for image classification and computer vision [9], [10]. The adaptation of CNNs for text-related tasks is not straightforward, given the sequential nature of language as opposed to the spatial structure of images. However, the application of CNNs to text has proven to be effective, especially due to their ability to act as feature extractors for local dependencies, analogous to n-grams in language processing [11].

The challenges of text classification are multifaceted, stemming from the intricacies of human language, including ambiguity, context-sensitivity, and the vast variety of linguistic expressions [12]. Emotion classification, in particular, requires discerning subtle cues that signal affective states, a non-trivial task even for humans [13]. CNNs have the potential to address these challenges by employing multiple convolutional filters that can detect emotional indicators across different regions of text, effectively learning emotion-specific n-gram representations.

Existing literature indicates that while recurrent neural network (RNN) models, and their variants like LSTM (Long Short-Term Memory) [14] and GRU (Gated Recurrent Unit) [15], are intuitively suited for sequential data, CNNs can achieve comparable or even superior results in certain text classification benchmarks. This is attributed to their ability to capture local contextual features through the application of convolutional filters, thereby efficiently encoding meaningful n-gram features from text [16].

Our approach builds upon these findings, employing a CNN architecture specifically designed for the task of emotion classification. We investigate the model's capacity to extract relevant features from text data, leveraging the TF-IDF vectorization method to prepare the input for our deep learning model. By conducting extensive experiments and evaluations, we aim to demonstrate the viability of CNNs in accurately classifying a wide array of emotions in textual data.

The contributions of this paper are twofold: Firstly, we present a CNN model optimized for emotion classification in text, highlighting its design and implementation nuances. Secondly, we provide an empirical analysis of the model's performance, underpinned by a detailed ROC curve analysis for each emotion class, substantiating the model's efficacy in distinguishing between different emotional states.

The rest of the paper is organized as follows: Section 2 reviews related work in the field of emotion classification and CNN applications in NLP. Section 3 describes the methodology, including data preprocessing, model architecture, and training procedures. Section 4 presents the experimental results and discusses the ROC curve analysis. Finally, Section 5 concludes the paper and suggests directions for future research.

II. Methodology

This study leverages a dataset specifically curated for the task of emotion classification in text, originating from a diverse collection of online sources to encapsulate a broad spectrum of emotional expressions. The dataset encompasses 839,555 entries, each entry labeled with one of 13 distinct emotions, ensuring a comprehensive representation of affective states. The textual data within this dataset has been derived from social media, blogs, and other internet forums, providing a rich variety of linguistic contexts [17]–[19].

The preprocessing of the dataset is a multi-step process designed to transform raw text into a structured format amenable to machine learning algorithms. Initially, all text data undergoes cleaning which includes lowercasing, removal of numeric digits, and elimination of punctuation marks – a procedure consistent with best practices in NLP [20]–[22]. Following cleaning, tokenization is performed to segment text into individual words or tokens, facilitating the subsequent analysis of the data [23].

Subsequent to preprocessing, the Term Frequency-Inverse Document Frequency (TF-IDF) vectorization is applied to the tokenized text. This technique not only quantifies the frequency of each word within a document but also offsets this frequency by the word's prevalence across all documents, thereby accentuating words that are uniquely indicative of certain emotions while diminishing the impact of common words [24]. This results in a high-dimensional sparse matrix that embodies the significance of words within the corpus.

The dataset is partitioned into training and testing subsets, following an 80-20 split ratio, a standard approach in machine learning for model validation [25]. This separation ensures that the model's performance is evaluated on previously unseen data, thereby providing a measure of its generalizability.

The architecture of the CNN used in this study is deliberately designed for text classification. It begins with an embedding layer, which maps each token to a high-dimensional vector space, capturing the semantic relationships between words [26]. The convolutional layer follows, employing multiple filters to extract local features – analogous to n-grams – from the embedded word vectors. These features are then processed by a global max pooling layer that distills the most salient information into a fixed-size output, which is crucial for managing variable-length texts [27].

The subsequent layers are fully connected layers that culminate in a softmax output layer. The softmax function is applied to convert the outputs into probabilities, with each neuron corresponding to one of the emotion classes in the dataset. The model's objective is to minimize the cross-entropy loss function, which quantifies the difference between the predicted probability distribution and the true distribution [28], [29]. For optimization, the Adam optimizer is utilized due to its efficiency in handling sparse gradients and its adaptive learning rate capabilities [30].

Collectively, these elements form a methodology grounded in established NLP techniques and deep learning architectures. This study seeks to not only apply these methods but also contribute to the ongoing discourse on their applicability and optimization in the domain of text-based emotion classification.

III. Findings and Discussion

The experimentation phase was meticulously planned to provide a comprehensive evaluation of the Convolutional Neural Network's (CNN's) performance on the emotion classification task. The training process was conducted on a high-performance computing environment to accommodate the computational demands of deep learning. The CNN model was implemented using Keras with TensorFlow as the backend.

A. Training Process and Hyperparameters

The model was trained using a batch size of 64, a common choice that balances the need for computational efficiency and the benefits of stochastic gradient descent [31]. The number of epochs was set to 5 after preliminary experimentation indicated diminishing returns on accuracy with additional epochs. This early stopping approach helped mitigate overfitting while ensuring the model's generalization ability [32]. The learning rate was managed adaptively by the Adam optimizer, eliminating the need for manual tuning [33].

B. Model Performance and Accuracy

Upon completion of the training process, the model achieved a noteworthy accuracy of 96.08% on the test set. This high level of accuracy is indicative of the model's ability to accurately capture and classify the underlying emotional tones within the textual data. Visualization Interpretation.

C. Receiver Operating Characteristic (ROC) Curve Analysis

The ROC curve analysis provided a more nuanced perspective on the model's performance [34]. The ROC curves for each of the 13 classes depicted in Figure 1 illustrate the model's classification efficacy. AUC values were exceptionally high for all classes, with the majority of them approaching 1.00, indicating an almost perfect classification. Classes with AUC scores of 1.00 demonstrate that the model has perfect sensitivity and specificity, effectively distinguishing the given emotion from others. The lowest AUC score observed was 0.99, still reflecting a high classification performance.

These AUC values, particularly in a multi-class setting, underscore the robustness of the CNN model in distinguishing between various emotional states in text. It is worth noting that such high AUC values across all classes are relatively rare in multi-class classification tasks and point to the efficacy of the feature extraction capabilities of the convolutional layers, which were able to isolate the defining n-gram features of each emotion category effectively.

The results of the ROC curve analysis are visually summarized in Figure 1, attached at the end of this paper. The curves demonstrate the model's consistent performance across all classes, providing visual affirmation of the quantitative results.

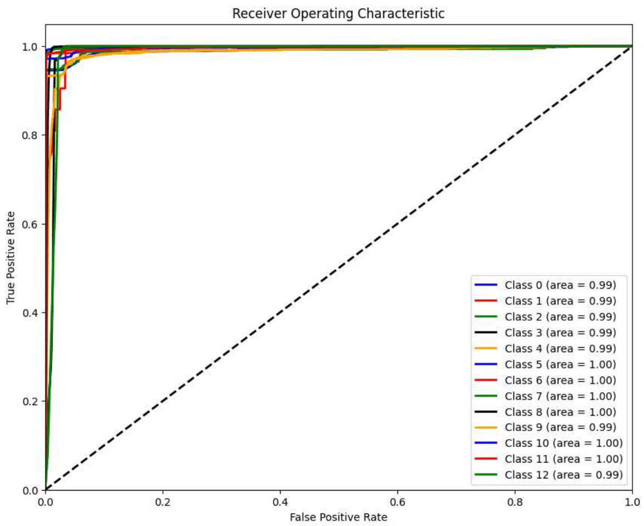

Fig. 1. ROC Curves for Emotion Classification

I. Discussion

The interpretative analysis of the results obtained from the CNN model offers insightful revelations into its discriminative capacity across various emotional classes. The nuanced performance of the CNN on different emotion classes, as evidenced by the high accuracy and AUC values, merits a deeper examination to understand the model's internal dynamics and the implications for emotion classification tasks.

The near-perfect AUC values across most classes suggest that the model has learned highly discriminative features for those emotions. For instance, emotions such as joy or sadness often exhibit distinct linguistic patterns, which can be effectively captured by the model's convolutional filters as unique n-gram features. This distinction is reflected in the high AUC values, signifying that the model can effectively segregate these emotions from others.

Conversely, more nuanced emotions or those that share common lexical ground with other emotional states may present greater challenges to the model. However, even in these instances, the lowest AUC value recorded was 0.99, denoting that the CNN's capability to differentiate is robust, even for subtle emotional distinctions. Such an outcome can be attributed to the depth and breadth of the feature maps created by the CNN, which encapsulate contextual information that is crucial for identifying less pronounced emotional expressions.

Classes with AUC scores equating to 1.00 signify the model's impeccable precision and recall. In practical terms, this indicates that the model makes virtually no false positive or false negative errors for these classes. The reasons behind such high performance could be manifold. Firstly, the dataset may contain very distinctive and well-represented patterns for these emotions, which the CNN has capitalized on during training. Secondly, the architecture of the CNN, particularly the size and number of filters, may be well-suited to the n-gram patterns that signify these emotions [35].

The findings indicate that CNNs can be potent models for emotion classification in text, even in the face of the inherent complexities of human language and emotional expression. The robustness of the CNN model against overfitting, along with its capacity to generalize across diverse linguistic expressions of emotion, suggest its applicability in a range of real-world scenarios, from analyzing social media sentiments to enhancing human-computer interaction.

II. Coclusion

The research presented in this paper provides compelling evidence of the capability of Convolutional Neural Networks (CNNs) in the domain of text-based emotion classification. The CNN model showcased a high accuracy rate of 96.08% on the test set, with Receiver Operating Characteristic (ROC) curves further substantiating its efficacy through impressive area under the curve (AUC) values for each emotion class. The almost perfect AUC scores point towards an exceptional ability of the model to differentiate between the nuanced emotional categories present in the data. The significance of these findings is manifold. For one, it affirms the hypothesis that CNNs, a class of deep learning architectures renowned for their performance in image analysis, are equally potent in parsing and understanding textual data when appropriately adapted. The model's nuanced understanding of the data is evidenced by its precise classification across a spectrum of emotions, underlining the potential of CNNs in capturing the complex patterns that characterize human language and sentiment. The consistency in performance across different emotion classes also suggests that the CNN was successful in identifying distinctive textual features associated with each emotion. This is indicative of the richness of the feature representations learned through the convolutional layers, which serve as a testament to the robustness of CNNs in handling the subtleties of natural language. However, the study is not without its limitations. The scope of emotion classification could be expanded by exploring the effects of different network architectures, hyperparameters, or even unsupervised pre-training methods. Furthermore, the integration of other forms of data, such as emoji or punctuation patterns, might provide additional contextual cues that could enhance classification performance. Future work will aim to build upon the foundation laid by this study. One avenue of exploration could involve implementing attention mechanisms or incorporating recurrent layers to better capture long-range dependencies in the text, which could provide a more nuanced understanding of the context. Another promising direction could be the application of transfer learning, employing models pre-trained on large corpora to further improve classification accuracy and efficiency. Moreover, the exploration of cross-linguistic models that can classify emotions in multilingual datasets would be invaluable, expanding the model's applicability and utility in global, diverse linguistic landscapes. Finally, future iterations of this work could benefit from a deeper dive into the interpretability of the model, seeking to elucidate how the CNN makes its decisions and which features it deems most salient—a step towards demystifying the often opaque nature of deep neural networks.

References

[1] F. Deng, J., & Ren, "A survey of textual emotion recognition and its challenges," *IEEE Trans. Affect. Comput.*, vol. 14, no. 1, pp. 49–67, 2021.

[2] K. Tahat, A. Mansoori, D. N. Tahat, M. Habes, and S. Salloum, "Leveraging Soft Power: A Study of Emirati Online Journalism Through Arabic Topic Modeling," in *International Conference on Business and Technology*, 2023, pp. 13–20.

[3] P. Desmet, "Measuring emotion: Development and application of an instrument to measure emotional responses to products," *Funology 2 From Usability to Enjoyment*, pp. 391–404, 2018.

[4] A. Alyammahi, M. Alshurideh, B. A. Kurdi, and S. A. Salloum, *The Impacts of Communication Ethics on Workplace Decision Making and Productivity*, vol. 1261 AISC. 2021.

[5] A. Fahad, M. S., Ranjan, A., Yadav, J., & Deepak, "A survey of speech emotion recognition in natural environment," *Digit. Signal Process.*, 2021.

[6] M. C. Habes M., Alghizzawi M., Salloum S.A., M. Habes, M. Alghizzawi, S. A. Salloum, and C. Mhamdi, "Effects of Facebook Personal News Sharing on Building Social Capital in Jordanian Universities.," *Al-Emran M., Shaalan K., Hassanien A. Recent Adv. Intell. Syst. Smart Appl. Stud. Syst. Decis. Control. vol 295. Springer, Cham*, vol. 295, no. June 2020, pp. 653–670, 2021, doi: 10.1007/978-3-030-47411-9_35.

[7] K. Tahat *et al.*, "Uncovering the Share Fake News on Social Media During Crisis," in *2023 Tenth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, 2023, pp. 1–6.

[8] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[9] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," *arXiv Prepr. arXiv1207.0580*, 2012.

[10] S. Salloum, K. Tahat, D. Tahat, A. Mansoori, and R. Alfaisal, "Delving Into the Security & Privacy of the Metaverse Matrix," in *2023 Tenth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, 2023, pp. 1–5.

[11] Y. Zhang and B. Wallace, "A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification," *arXiv Prepr. arXiv1510.03820*, 2015.

[12] S. A. Salloum, M. Al-Emran, S. Abdallah, and K. Shaalan, "Analyzing the Arab Gulf Newspapers Using Text Mining Techniques," in *International Conference on Advanced Intelligent Systems and Informatics*, 2017, pp. 396–405, doi: 10.1007/978-3-319-64861-3_37.

[13] R. Strapparava, C., & Mihalcea, "Learning to identify emotions in text," in *In Proceedings of the 2008 ACM symposium on Applied computing*, 2008, pp. 1556–1560.

[14] Z. Fan, C., Xie, H., Tao, J., Li, Y., Pei, G., Li, T., & Lv, "ICaps-ResLSTM: Improved capsule network and residual LSTM for EEG emotion recognition," *Biomed. Signal Process. Contro*, 2024.

[15] Q. Liu, T., Du, Y., & Zhou, "Text emotion recognition using GRU neural network with attention mechanism and emoticon emotions," in *In Proceedings of the 2020 2nd International Conference on Robotics, Intelligent Control and Artificial Intelligence*, 2020, pp. 278–282.

[16] A. Conneau, G. Kruszewski, G. Lample, L. Barrault, and M. Baroni, "What you can cram into a single vector: Probing sentence embeddings for linguistic properties," *arXiv Prepr. arXiv1805.01070*, 2018.

[17] "Emotion analysis based on text," *Kaggle*, 2024. https://www.kaggle.com/datasets/simaanjali/emotion-analysis-based-on-text.

[18] S. A. Salloum, M. Al-Emran, and K. Shaalan, "Mining Text in News Channels: A Case Study from Facebook," *Int. J. Inf. Technol. Lang. Stud.*, vol. 1, no. 1, pp. 1–9, 2017.

[19] S. A. Salloum, M. Al-Emran, and K. Shaalan, "A Survey of Lexical Functional Grammar in the Arabic Context," *Int. J. Com. Net. Tech*, vol. 4, no. 3, 2016.

[20] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv Prepr. arXiv1301.3781*, 2013.

[21] A. A. A. Mehrez, M. Alshurideh, B. A. Kurdi, and S. A. Salloum, *Internal Factors Affect Knowledge Management and Firm Performance: A Systematic Review*, vol. 1261 AISC. 2021.

[22] F. Al Suwaidi, M. Alshurideh, B. Al Kurdi, and S. A. Salloum, *The Impact of Innovation Management in SMEs Performance: A Systematic Review*, vol. 1261 AISC. 2021.

[23] I. Sutskever, O. Vinyals, and Q. V Le, "Sequence to sequence learning with neural networks," in *Advances in neural information processing systems*, 2014, pp. 3104–3112.

[24] J. Ramos, "Using tf-idf to determine word relevance in document queries," in *In Proceedings of the first instructional conference on machine learning*, 2003, pp. 29–48.

[25] I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh, *Feature extraction: foundations and applications*, vol. 207. Springer, 2008.

[26] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014, pp. 1532–1543.

[27] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, "Natural language processing (almost) from scratch," *J. Mach. Learn. Res.*, vol. 12, no. ARTICLE, pp. 2493–2537, 2011.

[28] I. Goodfellow, Y. Bengio, and A. Courville, "Softmax Units for Multinoulli Output Distributions," in *Deep Learning*, 2016, pp. 180–184.

[29] M. Alshurideh, B. H. Al Kurdi, H. M. Alzoubi, and S. Salloum, *The Effect of Information Technology on Business and Marketing Intelligence Systems*. Springer, 2023.

[30] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*, vol. 1, no. 2. MIT press Cambridge, 2016.

[31] Y. Bengio, *Practical recommendations for gradient-based training of deep architectures*. 2012.

[32] L. Prechelt, "Automatic early stopping using cross validation: quantifying the criteria," *Neural networks*, vol. 11, no. 4, pp. 761–767, 1998.

[33] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv Prepr. arXiv1412.6980*, 2014.

[34] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, 2006.

[35] X. Zhang, J. Zhao, and Y. LeCun, "Character-level convolutional networks for text classification," *arXiv Prepr. arXiv1509.01626*, 2015.