



Research Article

Augmentation of Contextualized Concatenated Word Representation and Dilated Convolution Neural Network for Sentiment Analysis

Fazeel Abid,^{1,2} Ikram Ud Din³, Ahmad Almogren⁴, Hasan Ali Khattak⁵, and Mirza Waqar Baig⁶

¹Department of Information Systems, School of Business and Economics, University of Management and Technology, Lahore, Pakistan

²School of Information Science and Technology, Northwest University, Xi'an, China

³Department of Information Technology, The University of Haripur, Pakistan

⁴Department of Computer Science, College of Computer and Information Sciences, King Saud University, Riyadh 11633, Saudi Arabia

⁵School of Electrical Engineering and Computer Science, National University of Sciences and Technology (NUST), Islamabad, Pakistan

⁶Department of Electrical Engineering, FAST National University of Computer and Emerging Sciences, CFD Campus, Faisalabad, Pakistan

Correspondence should be addressed to Ikram Ud Din; ikramuddin205@yahoo.com, Ahmad Almogren; ahalmogren@ksu.edu.pk, and Hasan Ali Khattak; hasan.alikhattak@seecs.edu.pk

Received 13 October 2021; Revised 26 October 2021; Accepted 29 October 2021; Published 25 November 2021

Academic Editor: Shalli Rani

Copyright © 2021 Fazeel Abid et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Deep learning-based methodologies are significant to perform sentiment analysis on social media data. The valuable insights of social media data through sentiment analysis can be employed to develop intelligent applications. Among many networks, convolution neural networks (CNNs) are widely used in many conventional text classification tasks and perform a significant role. However, to capture long-term contextual information and address the detail loss problem, CNNs require stacking multiple convolutional layers. Also, the stacking of convolutional layers has issues requiring massive computations and the tuning of additional parameters. To solve these problems, in this paper, a contextualized concatenated word representation (CCWRs) is initialized from social media data based on text which is essential to misspelled and out of vocabulary words (OOV). In CCWRs, different word representation models, for example, Word2Vec, its optimized version FastText and Global Vectors, and GloVe, collectively create contextualized representations upon the sequence of input. Second, a three-layered dilated convolutional neural network (3D-CNN) is proposed that places dilated convolution kernels instead of conventional CNN kernels. Incorporating the extension in the receptive field's size successfully solves the detail loss problem and achieves long-term context information with different dilation rates. Experiments on datasets demonstrate that the proposed framework achieves reliable results with the selection of numerous hyperparameter tuning and configurations for improved optimization leads to reduced computational resources and reliable accuracy.

1. Introduction

In topical years, progress towards intelligent applications showed excellent technological developments through social media data analytics [1]. These advancements are regulated

mainly and decently using social networks like Twitter, Facebook, and Instagram [2]. These social networks now transformed into a potential origin for mining social information to prevail over people's sentiments. The enormous opinions over social media comprise simple word sentences

and hold helpful information in several aspects. Consequently, social media data can be engaged to determine valued insights. Development in the pattern of social media data mining algorithms must focus on textual data. Sentiment analysis based on social media data is a rapidly evolving field to understand people's opinions, attitudes, and behaviors. An intelligent application can benefit social media sentiment analysis as these attitudes, feelings, and reactions can be correlated to the disasters, epidemic situations, government policies, and people perception, which is a substantial source of assessing the polarity: positive, negative, and neutral.

These applications concentrate on improvements in multiple aspects such as technological and legislative measures with social media data mining. It could also raise people's perspectives and cognition and endue them to acquire a viable environment [3]. The social media data on social networks consider a principal source of reviews against the events, disasters, and current epidemic situations though the challenge is a sizable social data that concerns efficient and scalable techniques to work on noise. The requirement for cleaning the noisy data requires automatic techniques for the classification of worthy information. It also goes through various issues, such as sentences written in short length, notations, and typos mistakes. In this regard, semantic exploration can adapt excessive social media data for syntactic regularities towards advancements through incorporation and scalability [4]. Thus, social media data mining algorithms consider gathering and handling social data effectively on Instagram, Facebook, and Twitter. We decide on twitter, which incorporates various posts and comprises 280 characters [5] for sentiment analysis.

Nowadays, machine learning methods are leveraged to augment services by mining social media data [6]. Among numerous methods of machine learning towards the classification of sentiments, Naive Bayes (NB) is exploited for topic detection [7], sentiment analysis [8], recommendation systems [9], and spam detection [10]. Further, a support vector machine that is a preferred technique in social media data has been applied [11]. However, due to the varying size of sequence composition, the stated methods are problematic towards extracting the features, which is significant. A sub-field of machine learning stated that deep learning incorporates neural architectures to extract expeditious high-level features while considering social media data classification. Also, techniques based on neural networks are increasingly utilized to solve the problems associated with supervised as well as unsupervised learning [12–14].

Deep learning methodologies assure researchers that the use of neural networks empowers extracting features devoid of involving the complex engineering of features [15, 16]. Feature extraction and classification are carried out through a sequence of words by multiplying with related weight as a one-hot vector or matrix [17]. The succession of a respective word is interpreted by way of continuous vector space initializing to neural architecture using several layers for prediction. This impacts the learning set to increase the classification evaluation metrics such as accuracy defined in [18]. Among neural architectures, convolutional neural

networks have attained adequate results to classify sentences obtained from social media [19, 20]. Multiple distributed word representations such as Word2Vec [21], GloVe [22], and FastText [23] can learn through mapping the words upon lower dimensions. A technique for extracting features using handcrafted features to classify sentences featuring convolutional neural networks introduced in [24] cannot hold long-term dependencies. Relatively, the employment of the CNN variant as dilated convolution removes the consequences containing information loss owing to traditional approaches of down sampling in conventional pooling operations and the stride convolution. Additionally, it scales receptive fields significantly, devoid of more parameters that make dilated convolution feasible to hold long-term dependencies and semantics.

Though many research works inadequately coupled relations regarding social media data ought to be intensified, this study modifies a deep neural network technique that automatically specifies social media data engaging a dilated convolutional neural network architecture in the parallel mechanism. From our best evaluation, a parallel mechanism in a dilated convolutional neural network can efficiently predict appropriate information by learning features from a contextualized concatenated word representational model using different embeddings. Beyond the developments, various hyperparameters employing dilated convolutional neural networks to analyze social media data sentiments are reasoned. This paper sets up a new approach for sentiment analysis and is utilizable to improve many services. The following are the contributions of this work:

- (i) Contextualized concatenated word representational (CCWRs) model is utilized to get classifier's improved exhibition features compared with many state-of-the-art techniques
- (ii) A parallel mechanism in three dilated convolution pooling layers featured different dilation rates, and two fully connected layers in a novel approach are considered
- (iii) Lastly, the work undertakes a deep learning approach using multiple parameters and hyperparameters to offer intelligent applications using Twitter data for sentiment analysis to enhance people's behavior

The rest of the paper is organized as follows. Section 2 continued with related work. The proposed framework for sentiment analysis is accessible in Section 3. Section 4 covers experimental setups and results, while the discussion is in Section 5. Finally, the paper is concluded in Section 6.

2. Related Work

The continuous social media data maturation has incited an advanced degree in scientific and sustainable smart urban explorations. Plenty of tasks have been performed over sustainability toward smart applications by social media networks [25, 26]. Still, the collective signification is not yet

entirely considered and admired [27]. Social media users now appreciate the accessibility and necessity of smart services that imply smart applications' requirements by concentrating on social media contexts [28, 29]. These smart applications activate the combination of social media networks and a smart environment for the social user's opinions, and prospects possess a sound impact, as explained in [30, 31]. By concentrating on social media networks with associated information, like hashtags, time, location, and name, the present work intends to explore how these networks fundamentally contribute to elevating the importance of smart applications. Therefore, social media network users regarded as smart application sensors together with associated metainformation can be utilized in many research works as described by [32–34].

Moreover, the data composed by social users tend to be syntactically unique and ubiquitous using smartphones, which can be appropriately collected and analyzed the information in a short textual framework [35]. Numerous multiple perspectives, from event detection to disease tracking and monitoring employing small textual content on social media, have been proposed in [35–37]. This short textual content of social media networks such as Twitter is semantically essential and extensively utilized in heterogeneous applications related to text classification [38, 39].

A methodology using features (metalevel) considering emotions from Twitter data for the polarity classification has been offered [40]. Many methodologies of manual tagging data gathered from Twitter with metainformation similar to location and social user for training based on conditional random fields as presented in [41, 42]. Similarly, multiple techniques categorize different topics like joy, fear, anger, love, and surprise by tagging on Twitter proposed in [43, 44], to accomplish emotional analysis. To originate smart applications, Twitter can identify numerous aspects such as people inferring and trends. However, these aspects are bound toward noise from nonassociative contents, which are crucial, as clarified in [45]. Therefore, a filter should be considered to attain adequate associative information on mentions, URLs, slang words, and numbers. A set of features as the evaluation metrics rely on features class.

Machine and deep learning-centered techniques employing social media data must be essential to mine valuable insights as presented in [46, 47]. The mining process, such as social media data, intends to assess people's opinions in many prospects, such as gathering data linked to the user and observing interactions [47]. Overall representation starts from unstructured to structured data; deep learning performs better than machine learning methods, which are time-consuming and require complex manual feature engineering processes. These positive aspects in deep learning methods to mine opinions from enormous social networks in streaming, multimedia, and textual framework provoked researchers in numerous works [48–52]. Therefore, extracting opinions from social networks such as Twitter in text form using unsupervised learning is significantly regarded for appropriate representation in this work.

Among many neural networks, traditional convolutional and recurrent neural networks have prominently accom-

plished higher outcomes from social networks for capturing long-term dependencies and extracting the opinions, such as sentiment analysis. An optimized version of recurrent neural network, long-short-term memory utilized to improve the semantics, is proposed in [53], but the training of mentioned version was computationally difficult. To extract syntactic features and perform faster training for text classification based on social media, convolutional neural networks testified to be more suitable [54, 55]. Further, a convolutional neural network as a joint trained task can substantially extract features as well as classification [56]. Many researchers are progressively employing convolutional in recent works with pooling layers for morphological modeling [57, 58].

Similarly, to cope with contextual data in character and sentence-level, two convolutional layers of deep architecture for the classification of short texts are offered in [59]. However, conventional convolutional architecture requires multiple layers stacking with the length of text. An improved form referred to as dilated convolution neural network; comparatively, a more sensible choice capable of the increased size of receptive field size adequately overcomes the issues and utilized in many works [60–62].

Distributed word representation in deep learning transforms words into a continuous vector; likewise, pretrained learning representation makes an essential impact while classifying the social data for sentiment analysis. It is also observed that social data classification augments the learning toward syntactical, phonological, and sentimental information; some of the works attempt to combine pretrained vectors. But the syntactical and phonological issues are demanding relationships between the words for the sufficient and actual classification as explained in [63–65], although the concept of combining varied pretrained representations is of significant results concerning different channels towards the classification of multiple sentences as presented in [66]. However, the dimensions of combined representations should be the same, restricting the scope and usage of pretrained representations due to multiple dimensions.

Furthermore, CNN has been employed to identify the actual word events from sentence-level social data by considering position along with entity explained in [67]. A more in-depth work focused on social media data pursuance of sentence-level classification confirms the ubiquity of multiple opinions or events in [68]. In this work, a dynamic multipooling layer is introduced to extract opinions about events for improved information. Although CNNs have been in continuous considerations among researchers, long-term dependencies regarding semantic features toward input sequence remain challenging. Also, it has been observed that CNNs tend to depend on stacking various layers together with the convolution-pooling to adapt long-term semantic dependencies. From the best of our knowledge, a dilated convolution network that adjusts dimension and encompasses addition in the receptive field's size devoid of loss of detailed information and the problem of long contextual and semantic dependencies is addressed effectively through varied dilation rates.

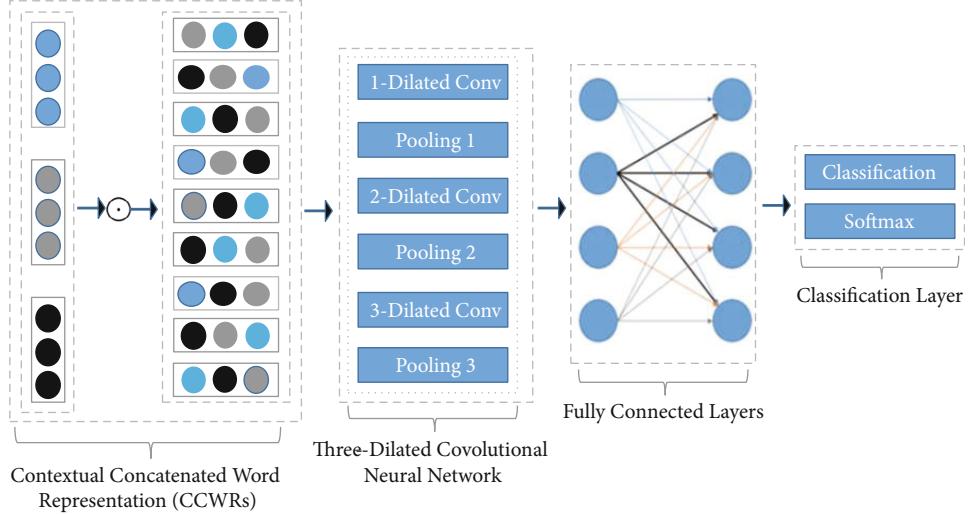


FIGURE 1: Structure 3D-CNN with 2 fully connected layers followed by contextualized concatenated word representations.

3. The Proposed Architecture for Sentiment Analysis Based on CCWRs and 3D-CNN

This section contributes a theoretical model regarding a potential approach for social media data considering Twitter for sentiment analysis based on COVID's perspective. Initially, different word representational models are concatenated, referred to as contextualized concatenated word representation (CCWRs). Second, the 3D-CNN architecture in this work is made by utilizing three dilated convolution kernels and two fully connected layers to seizure long-term contextual dependencies concerning semantic features. We utilized multiple dimensional convolution processes to manage additional complexities toward enhanced performance by initializing the words to the 3D-CNN as a matrix to extract sufficient features through corresponding weights. Our proposed architecture has subsequent portions: concatenated word representations, three dilated convolution-pooling layers, two fully connected layers, and Softmax (as shown in Figure 1).

3.1. Contextualized Concatenated Word Representations (CCWRs). To represent the words as dense vectors, word representation models are regarded as essentials for feature extraction. These word representations are effective and result in advances toward the execution of social media data sentiment analysis as described in [69–71]. Many recent works have considered improved word and feature representations by way of different word embedding models [65, 72, 73]. Though these models diversify in architecture and pretraining, they still encode the input according to the surroundings. Words are represented utilizing only a single pretrained language model.

Further, these representations are unfeasible as a result of slow training and evaluation. Using pretrained models trained on multiple datasets exploits the biasness in different datasets, leading to numerous representations associated with the same word. On the other hand, the concatenation

of multiple word representational model can produce better representations devoid of computational complexities compared to a single model nearer to contextualized embeddings.

In this work, different pretrained word representational models such as Word2vec [21], fastText [23], and GloVe [22] are concatenated to deal with sentiments regarding contextualized and semantic information through the weighted mechanism. We leverage multiple word representations to produce a single table for every pretrained model in which the token of related input is embedded into a single vector space. Then, the subsequent vectors tend to be concatenated toward a single vector. Such weighted concatenation sufficiently upgrades the semantics and can handle the most recent problems identified with the misspelled and out of vocabulary words. The process of concatenation using associated weights assists in exploring better representations and functionally helps for sentence encoding in pursuance of feature selection. We utilize GloVe trained on Twitter, having 2 billion tweets, 27 billion tokens, and 1.2 million vocabulary, Word2vec 30 million tweets, and google news. In contrast, fastText on 1-million-word vectors, 16 billion tokens with subword information on Wikipedia, UMBC web-based corpus, and <http://statmt.org> news dataset considering dimensions range from 100 to 200.

We discard words that lie less than ten times and convert the characters to lowercase, and the most acceptable size corresponding to context window size selection is 5. In the training of CCWRs, we proportionately dropped the learning rate with the improvement in training. It has been observed that early regarded text does have an overall impact on the precession of the model. In this work, we train the concatenated representations on multiple datasets associated with different anomalies in the world. The first dataset is congregated by the use of TAGS and streaming API as described in [74]. The variety of keywords seemed to be evolving incessantly over social media. However, to stream the tweets of contextual perspective, the rational filtering keywords of our work exhibited in the table accumulated

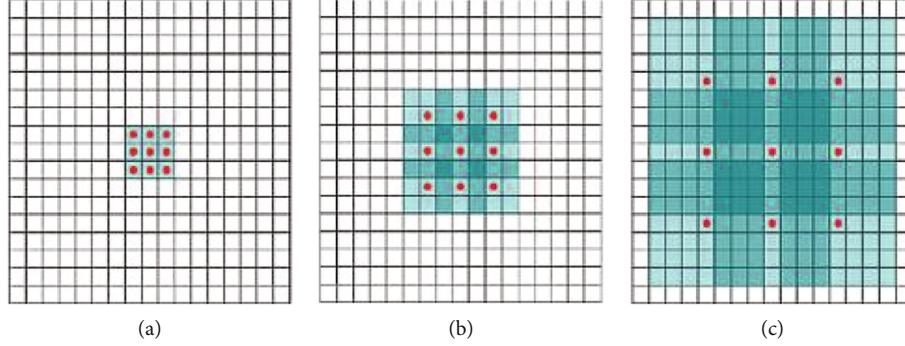


FIGURE 2: The process of stacking and its effect of dilated convolution kernel [60].

from September 2020 to March 2021 are still in use. The other dataset considers the anomalies in the real world based on Twitter as utilized in [75].

3.2. Three Dilated Convolutional Neural Network (3D-CNN). In sentiment analysis based on the textual framework, the conventional CNN, due to the pooling layer, some significant text features are missed during calculation, ending in the adverse effect in overall network precision. Similarly, to acquire significant features, CNN architecture is deepened by stacking more layers, which concerns more parameters and additional computational resources. Also, the backpropagation of gradient may lead to vanishing gradient while increasing layers in network, causing performance to reduce significantly. Further, the limited size of convolutional kernels causes classical CNN only to hold short-term dependencies of text. To handle stated issues, a process that places zeros into the primary convolution kernel formed the dilated convolution kernel as introduced in [60]. This placement by way of intensifying receptive field size enables capturing more information and uplifting network's entire performance, for example, Figure 2 by which conventional convolution kernel of 3×3 size, a point upon 0 weight placed on each point in a matrix in Figure 2(a) and then develop in Figure 2(b); equally, Figure 2(c) exhibits as receptive field size $3 \times 3, 7 \times 7, 15 \times 15$ concerning convolution kernel.

The receptive field size tends to increase as the placed holes increase. However, it is observed that the parameters are the same as shown in Figures 2(a)–2(c). The dilated convolution kernel seems to process the text enabling the convolution kernel to obtain additional information without additional computational resources. Therefore, the increase in receptive field size is essential for many tasks such as prediction and classification. A three dilated convolutional layers model is presented in Figure 3 showing the significant rise in dilation rate at each layer. The model with a specific feature map as (F_1, F_2, F_3, F_4) of dilation convolution (1, 2, 3, 4) together with the receptive field size of each element as $((3 \times 3), (2^3 - 1) \times (2^3 - 1), (2^4 - 1) \times (2^4 - 1))$ where receptive field size of each element in F_{i+1} by way of $(2^{i+2} - 1) \times (2^{i+2} - 1)$ can be seen.

To maximize the performance of the dilated CNN over the traditional CNN model, a novel architecture 3D-CNN following CCWRs containing three dilated convolution

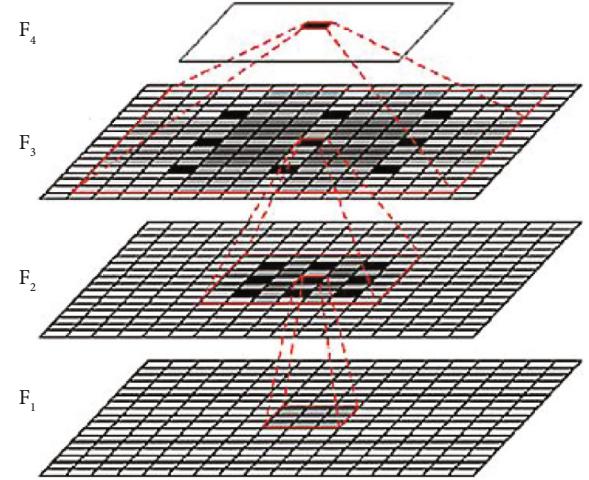


FIGURE 3: Three dilated convolution layers [76].

and pooling layers via two fully connected layers is proposed. The increase in the receptive field size extracts sufficient linguistics and contextual information without affecting and extending dimensions and parameters. This implementation efficiently increases the convolution kernel considering multiple scales with the aid of dilated operation while applying distinctive dilation rates as described in [77] and shown in Figure 4.

The choice of these dilation rates is significant when designing the structure of 3D-CNN as mentioned in [21] depends upon:

$$D_i = \max [D_{i+1} - 2d_i, D_{i+1} - 2(D_{i+1} - d_i), d_i]. \quad (1)$$

Here, $i = 1, 2, \dots, n$; d_i as dilation rate toward the i^{th} layer, whereas D_i being foremost d_i in the prescribed layer. Figure 5(a) has three dilated convolution kernels utilizing 3×3 as size with $d_i = 2$, similarly, Figure 5(b) using $d_i = 1, 2$, and 3.

The considered dilated convolution kernels (1, 2, 4) over several dilation rates (1, 2, 4) ensure the extraction of each semantics and not obstinate contextual information while extracting the feature maps. Our model significantly extracts the semantics and contextual information for sentiment

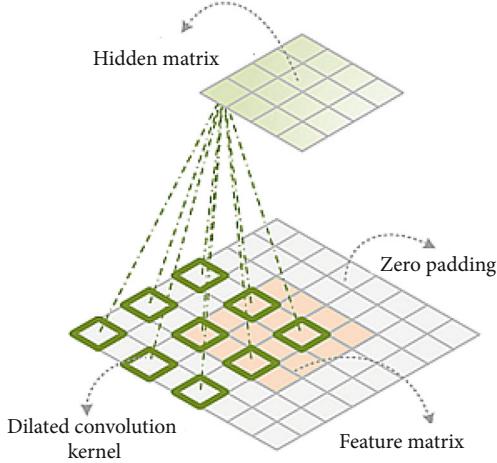


FIGURE 4: The dilated convolution operations over multiple dilated rates ($2 * 2$) [77].

words and sentences by considering multiple dilation rates. The three dilated convolution-pooling layer calculation for the extraction of long semantics and contextual information following CCWRs is presented in the following articulations:

$$D_c^1 = \text{RReLU}(F *_{d_1} X_t^x), \quad (2)$$

$$D_c^i = \text{RReLU}(F *_{d_i} D_c^{i-1}), \quad (3)$$

$$D_c^L = \text{RReLU}(F *_{d_L} D_c^{L-1}), \quad (4)$$

where the dilated convolution is $F *_{d_i}$ in the particular layer.

4. Experiments

This section initially incorporates the datasets, experimentation, and analysis of the results via different methodologies compared with multiple datasets. The main goal is to evaluate the proposed novel technique; this work presents an appropriate classifier including three dilated convolutional layers accompanied through CCWRs.

4.1. Datasets. To precisely evaluate, we test the proposed model on two datasets towards suitability, adaptability, and reliability. The first dataset encompasses various 27 events in real-world life, such as disasters, emergencies, and incidents that are publicly accessible [78]. The second dataset is congregated by streaming and TAGS [79]. These API are considered as interface Twitter search interface, which utilizes the keywords and terms specified by the user. The user places a query, and TAGS retains the results through a free google sheet and offers setup tags to update the sheet whenever needed. The keywords selected for tweets gathered from September 2020 to March 2021 are mentioned in Table 1 as search terms accumulate to 18920. Social media data like Twitter carry a lot of noise, such as numbers, URL, and user mentions, to normalize the data and handle redundancy; some preprocessing techniques, tokenization, and lemmatization are considered described in [80]. Further, tweets fea-

turing only five words as well as the stop words decided to eliminate. A shuffling in both datasets is determined for reliable outcomes, applicability, and appropriate analysis since all datasets are equal for better performance [81]. Four sets from the shuffling of selected datasets are acquired in which each fuses a comparable amount of tweets alluded to ESD1, ESD2, ESD3, and ESD4 as equally shuffled data.

4.2. Experimental Setup. The experimentations accumulated using diverse datasets coupled with multiple word representations language models to provide contextual concatenated word representation by considering suitable parameters. The activation functions, optimization algorithm, training, mini-batch size, filter size, number of hidden layers, the receptive field size, and the number of epochs incorporated are presented in Table 1. To deal with the issue of vanishing gradient in training, the rectified linear unit (Relu) and hyperbolic tangent (Tanh) are taken into consideration, which generally sets the output and serves input to neurons in the subsequent layer as explained in [82]. For the regular distribution and to reduce overfitting, a variant of Relu, a randomized (RRelu), is also considered through which the parameters with regard to negative impacts are sampled randomly [83].

For training, optimization algorithms such as stochastic gradient descent SGD, a 0.01 learning rate, and a stochastic optimization ADAM learning rate of 0.001 are utilized. Further, to improve training performance, the root mean squared propagation (Rmprop) optimizer that calculates the gradients upon a fixed window is regarded. There is no conceptualization for a specific choice of neurons in hidden layers; similarly, the wrong choice of neurons results in underfitting and overfitting due to few or more neurons that ultimately influence model's training [84]. Keeping in mind the nature of this work, the different choices of neurons 150, 300, and 400 in hidden layers are adequate to evaluate.

Moreover, training in neural architectures is reasonable using minibatches to split the large datasets into smaller sets. The minibatch gradient descent is taken into account with batch sizes of 64, 128, and 256, respectively. On the other hand, considering 10, 15, and 20 widths of the model since model's width is determined by the choices of hidden layers that impact the entire complexity of the neural network architecture. Similarly, for generalization, epochs refer to the number of times a dataset tends to pass through the network and cause the model to under or overfit due to selected epochs number. The selected number of epochs continues to be 10 to 100 for best critical analysis by considering mentioned hyperparameters. Lastly, varied filter size incorporates $(2, 2, 3)$, $(2, 2, 4)$, $(2, 2, 5)$, as well as the dilation width is set to $(1, 2, 4)$.

4.3. Results and Analysis. We completed experimentations on numerous assessment metrics on equally shuffled datasets by way of baseline and proposed methods. Precession, recall, classical metric accuracy, and $f1$ -score to inspect the symmetry in recall and precision are considered to deal with the imbalanced data. A 16 core processor of 3000 MHZ and 32 GB RAM is used to accomplish all the experimentation. Additionally, ML library "TensorFlow," an open-sourced

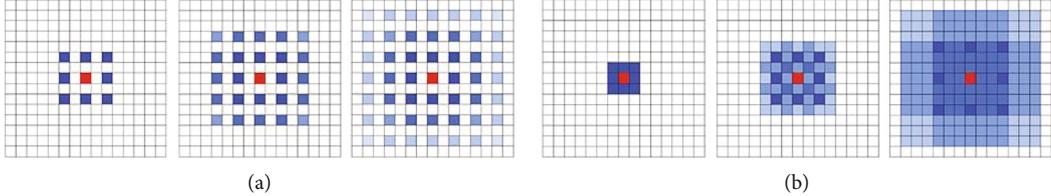


FIGURE 5: The stacking effect of dilated convolution kernels [21].

TABLE 1: Selected keywords for data selection and gathering.

| Period | Filtered keywords |
|------------------------------|---|
| September 2020 to March 2021 | #corona, #coronavirus, #first wave of Coronavirus, #firstwaveofCoronavirus, #pandemic, pandemic, #epidemic, epidemic, #handsanitizer, #covid, #covid19, #sarscov2, #ncov2019, #2019-ncov, #vaccination trials #lockdown, #quarantine, #Socialdistance, #Social Distance, #flattencurve, #flatteningcurve #working from home, #workfromhome, #second wave of Coronavirus, #secondwave of coronavirus, #huge number of deaths, #deadly virus, #deadly wave, #againlock down n95, #K95, #N95, #k95, #Govt Polices on Covid, #Govt relied on Covid. #Covid Vaccine, #Covid Vaccination, #Pfizer, #Astrazeneca, #Astra Zeneca, #side effects of vaccinations, #price of vaccination # availability of Pfizer, #chinese vaccine #sinopharm #new variant of Covid, #third wave # third wave of covid #more mutation #Moremutation, # relaxation in lockdown, #immunity in covid, #vaccination started |

[85], is involved in the training and comparing the proposed framework [85]. We settled the evaluation by including baseline models, which contained each model and pre-trained word vectors compared with the proposed framework, with each pretrained vector on similar architecture involving hyperparameters for the comparative analysis carried out.

The evaluation metrics for baseline models are accuracy together with *F1*-score. The most elevated accuracy accomplished in baseline models is 74.04% and *F1*-score 70.42% by utilizing FastText and 73.65% and *F1*-score 70.64% through GloVe ESD-1, which are presented in Tables 2 and 3 and displayed in Figures 4 and 5.

The accuracy achieved on the proposed model is mentioned in Tables 4, 5, 6, and 7, along with hyperparameter settings selected and shown in Table 8. The employment of hidden layers seems to have a significant impact on the improvement of network tuning. Astonishingly, the optimization algorithm Rmprop attains dependable accuracy of 77.92% on ESD-4 with a batch size of 256, 20 hidden layers, and 300 neurons using randomized rectified linear units (RRelu), whereas the selection of other parameters accomplished slightly closer accuracy 77.80% on ESD-3 with a batch size of 256, 15 number of hidden layers, and 150 neurons using rectified linear unit (Relu) through SGD and 76.84% on ESD-4 with a batch size of 128, 15 number of hidden layers, and 150 neurons using tanh through ADAM, respectively. Further, the batch sizes from 64, 128, and 256, in our architecture achieved the best accuracy on the equally shuffled dataset denying many works claiming the best performance with 2 to 32 batch size. The other evaluation metrics for handling imbalanced data are precision, *F1*-score, and recall (as shown in Tables 9, 10, 11, and 12).

TABLE 2: Accuracy on baseline models.

| Methods | ESD-1 | ESD-2 | ESD-3 | ESD-4 |
|-----------------|--------|--------|--------|--------|
| GloVe-3D-CNN | 73.65% | 71.98% | 70.91% | 69.89% |
| Word2Vec-3D-CNN | 71.52% | 69.82% | 72.47% | 70.93% |
| FastText-3D-CNN | 74.04% | 73.97% | 72.84% | 72.76% |

TABLE 3: *F1*-score on baseline methods.

| Methods | ESD-1 | ESD-2 | ESD-3 | ESD-4 |
|-----------------|--------|--------|--------|--------|
| GloVe-3D-CNN | 70.64% | 66.80% | 69.57% | 56.51% |
| Word2Vec-3D-CNN | 71.52% | 68.67% | 68.49% | 60.38% |
| FastText-3D-CNN | 70.42% | 69.25% | 71.57% | 67.2% |

The most efficient results in consideration of precession are 79.21% on ESD-1, and recall is 79.08%, while *f1*-score is 76.82%, shown in Figures 6, 7, 8, and 9, which reveals proposed architeture's significance.

5. Discussion

Deep learning-based methodologies promoted the significant availability of word representations models such as Word2Vev, GloVe, and Fasttext. This work investigates the quality of the different word representation models to perform social media sentiment analysis for intelligent applications. Our work referred to the collection, selection, and evaluation of multiple standard metrics and appropriate hyperparameters mentioned in Table 8. The foremost challenge during this work is related to dimensions of multiple word representation models by way of weighted

TABLE 4: Proposed method accuracy on ESD-1.

| Datasets | Optimization algorithm | Epochs | Activation function | Neurons | Hidden layers | Accuracy | Batch size |
|----------|------------------------|--------|---------------------|---------|---------------|----------|------------|
| ESD-1 | RMprop | 10-100 | RRelu | 300 | 20 | 77.34% | 256 |
| ESD-1 | SGD | 10-100 | Relu | 75 | 10 | 76.96% | 64 |
| ESD-1 | ADAM | 10-100 | Tanh | 150 | 15 | 76.35% | 128 |

TABLE 5: Proposed method accuracy on ESD-2.

| Datasets | Optimization algorithm | Epochs | Activation function | Neurons | Hidden layers | Accuracy | Batch size |
|----------|------------------------|--------|---------------------|---------|---------------|----------|------------|
| ESD-2 | SGD | 10-100 | Relu | 75 | 10 | 77.69% | 64 |
| ESD-2 | RMprop | 10-100 | RRelu | 300 | 20 | 75.91% | 256 |
| ESD-2 | ADAM | 10-100 | Tanh | 150 | 15 | 76.46% | 128 |

TABLE 6: Proposed method accuracy on ESD-3.

| Datasets | Optimization algorithm | Epochs | Activation function | Neurons | Hidden layers | Accuracy | Batch size |
|----------|------------------------|--------|---------------------|---------|---------------|----------|------------|
| ESD-3 | ADAM | 10-100 | RRelu | 300 | 20 | 76.77% | 64 |
| ESD-3 | RMprop | 10-100 | Tanh | 75 | 10 | 76.27% | 128 |
| ESD-3 | SGD | 10-100 | Relu | 150 | 15 | 77.80% | 256 |

TABLE 7: Proposed method accuracy on ESD-4.

| Datasets | Optimization algorithm | Epochs | Activation function | Neurons | Hidden layers | Accuracy | Batch size |
|----------|------------------------|--------|---------------------|---------|---------------|----------|------------|
| ESD-4 | ADAM | 10-100 | Tanh | 150 | 15 | 76.84% | 128 |
| ESD-4 | RMprop | 10-100 | RRelu | 300 | 20 | 77.92% | 256 |
| ESD-4 | SGD | 10-100 | Relu | 75 | 10 | 76.98% | 64 |

TABLE 8: Selected parameters and hyperparameters.

| Activation function | Optimization algorithm | Neurons in hidden layer | Minibatch | Hidden layers | Epochs | Dilation filter size | Dilation width |
|---------------------|------------------------|-------------------------|-----------|---------------|--------|----------------------|----------------|
| Relu | Rmprop | 150 | 64 | 10 | | 2,2,3 | |
| Tanh | ADAM | 300 | 128 | 15 | 100 | 2,2,4 | 1,2,4 |
| RRelu | SGD | 400 | 256 | 20 | | 2,2,5 | |

TABLE 9: Proposed method performance metrics on ESD-1.

| Datasets | F1-score | Recall | Precision |
|----------|----------|--------|-----------|
| ESD-1 | 76.22% | 73.67% | 79.21% |
| ESD-1 | 76.82% | 75.98% | 76.37% |
| ESD-1 | 76.59% | 76.28% | 76.91% |

TABLE 11: Proposed method performance metrics on ESD-3.

| Datasets | F1-score | Recall | Precision |
|----------|----------|--------|-----------|
| ESD-3 | 75.72% | 72.59% | 77.10% |
| ESD-3 | 75.28% | 76.64% | 76.39% |
| ESD-3 | 76.81% | 73.01% | 77.24% |

TABLE 10: Proposed method performance metrics on ESD-2.

| Datasets | F1-score | Recall | Precision |
|----------|----------|--------|-----------|
| ESD-2 | 74.26% | 71.90% | 76.50% |
| ESD-2 | 73.36% | 71.88% | 75.63% |
| ESD-2 | 73.61% | 72.41% | 77.46% |

TABLE 12: Proposed method performance metrics on ESD-4.

| Datasets | F1-score | Recall | Precision |
|----------|----------|--------|-----------|
| ESD-4 | 75.07% | 72.60% | 77.84% |
| ESD-4 | 76.38% | 76.64% | 75.93% |
| ESD-4 | 75.19% | 79.08% | 76.14% |

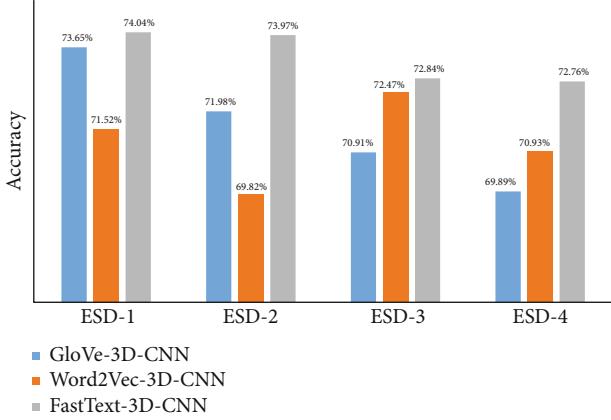


FIGURE 6: Baseline models accuracy on ESD-1, ESD-2, ESD-3, and ESD-4.

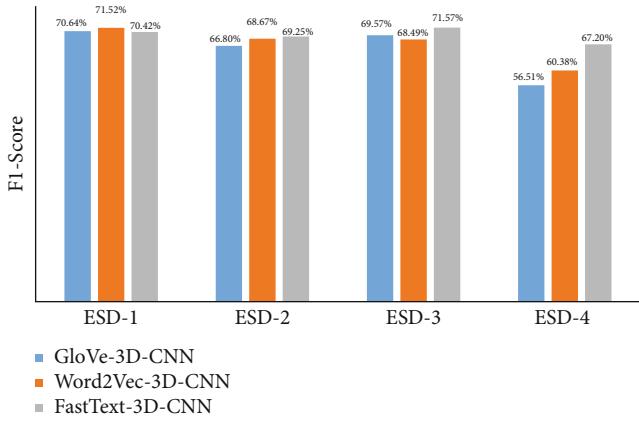


FIGURE 7: Baseline models on F1-score on ESD-1, ESD-2, ESD-3, and ESD-4.

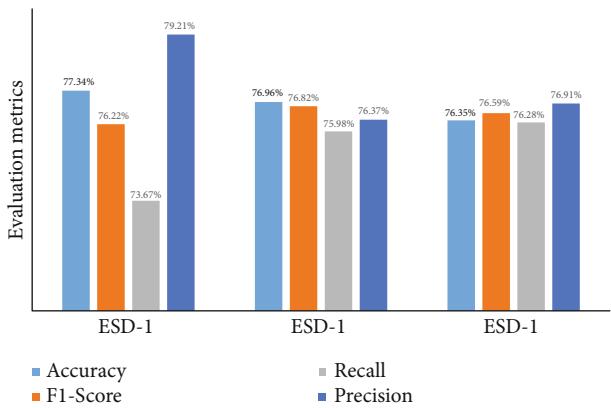


FIGURE 8: Evaluation metrics on proposed method on ESD-1.

concatenation to produce novel contextual concatenated word representation (CCWRs). The maturation of dilated convolution neural network architecture referred to as 3D-CNN is employed to increase the scale of receptive fields with different dilation rates to attain long-term contextual

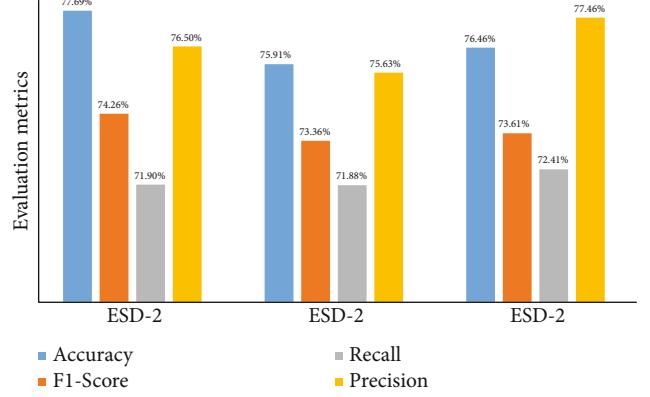


FIGURE 9: Evaluation metrics on proposed method on ESD-2.

regularities. The 3D-CNN architecture incorporates three dilated convolutional layers and a pair of fully connected layers. Following CCWRs, the processing of successive textual data and the computational time is spatially regulated by a succession of text [62].

Throughout this work as evident from Figures 10 and 11, it is observed that merely the stacking process of dilated convolution kernels effectively reduces training time and raises training accuracy to a definite level; however, not satisfactory enough to enhance the testing accuracy. This happens due to discontinuation between the dilated convolutions kernels, which captures minor information causes to neglect the constancy of information. Also, fixed-rate size during the extraction of feature maps, the big and little size information, cannot be considered simultaneously. These issues influence the training as well as testing accuracy of the fixed dilated convolutional model. In our work, novel CCWRs with 3D-CNN dilated varying dilation rates in the multiple layers utilize convolution operations series to capture complete information devoid of holes or missing. This successfully avoids information loss and the problem of testing accuracy using different dilated convolution kernels by increasing receptive field size.

By correlating the multiple distributed word representation model and contextual concatenated word representation model, we acknowledge that the development of CCWRs is significant despite including the small size of the corpus. Our experimentation provides the implementation of 3D-CNN in terms of important revelations such as (i) multiple word representation models by way of weighted concatenation for the generation of contextual representation along with two fully connected layers to classify social media data utilizing the linguistics regarding social media for intelligent applications and (ii) comparing and analyzing the optimization, preference, selection, tuning, and configuration of multiple parameters indicates the significant effect on the entire structure.

Nowadays, data available on social platforms, such as Twitter, is frequently used and has exceptional impacts on making intelligent and informed decisions marking which can be analyzed concerning people's opinions toward real-world events. Though many methodologies have been examined, it is still unable to mine out of vocabulary, misspelled,

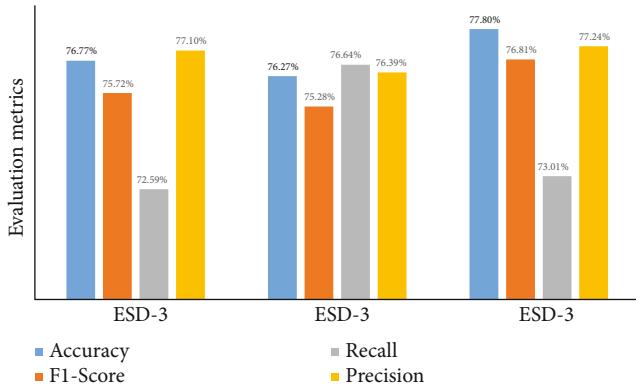


FIGURE 10: Evaluation metrics on proposed method on ESD-3.

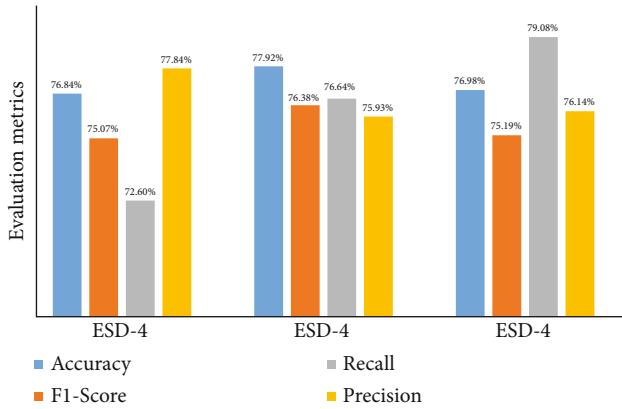


FIGURE 11: Evaluation metrics on proposed method on ESD-4.

and simple words to analyze social insights and the utilization of new neural architectures for intelligent applications. This work is also essential to social media textual data mining algorithm to consider real-word situations like disasters and current COVID-19 that entails well-timed effective techniques by observing people's impulses to assist the government in policy and strategic decisions.

Further, this paper is a significant source of accessibility of authentic, powerful, and evolving techniques concerning authorities necessary to consider the varying situation of the world with multiple variants of COVID-19. More in-depth, the idea can also extend to empower smart cities by contributing new methods through professionals by developing intelligent applications in epidemic situations towards the robustness of techniques and interpretations. To institutionalize intelligent applications regarded as an essential means, there is a need for propositions to use social media textual data mining algorithms in an intelligent environment involving a rapidly increasing social media textual data size. We can say that the development of the proactive, responsive, and cost-effective intelligent application will remain inadequate while performing without inheriting the significance of deep learning approaches and, more importantly, mining of insights of social media data.

6. Conclusion

The significance of social media data established an essential mean to realize people's attitudes to improve service. This paper uniquely formulates several hyperparameter tuning, selection, and configurations towards maximum model optimization on different valuation metrics. Proposing contextual concatenated word representations (CCWRs) trained on streamed social media data effectively surpasses various word representation models and overcomes out of vocabulary (OOV) words problem to some extent. Also, a novel proposition of three dilated convolution layers (3D-CNN) upon different dilated convolution kernels at each layer instead of stacking convolutional layers is utilized via a series of experimentations and verifications on multiple datasets. The proposed architecture as the augmentation of CCWRs and (3D-CNN) in the manner above accurately performs with many views such as avoiding loss of detailed, informative messages and capturing the long contextual information. However, it has been concluded that specific extensive training social media data can be helpful to extend evaluation metrics. Further, in our method, the imbalanced training data and subject-based collection of social media data from Twitter through relevant keywords is still a challenge that can be dilated in future work.

Data Availability

The data used to support the findings of this study are included in the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the King Saud University, Riyadh, Saudi Arabia, through researchers supporting project number RSP-2021/184.

References

- [1] P. B. Anand and J. Navío-Marco, "Governance and economics of smart cities: opportunities and challenges," *Telecommunications Policy*, vol. 42, no. 10, pp. 795–799, 2018.
- [2] A. Mainka, S. Hartmann, W. G. Stock, and I. Peters, "Looking for friends and followers: a global investigation of governmental social media use," *Transforming Government: People, Process and Policy*, vol. 9, no. 2, pp. 237–254, 2015.
- [3] C. T. Yin, Z. Xiong, H. Chen, J. Y. Wang, D. Cooper, and B. David, "A literature survey on smart cities," *Science China Information Sciences*, vol. 58, no. 10, pp. 1–18, 2015.
- [4] J. F. F. Pereira, "Social media text processing and semantic analysis for smart cities," 2017, <http://arxiv.org/abs/1709.03406>.
- [5] R. Passonneau, "Sentiment analysis of Twitter data," in *Proceedings of the Workshop on Language in Social Media (LSM 2011)*, pp. 30–38, Portland, Oregon, 2011.

- [6] J. Chin, V. Callaghan, and I. Lam, "Understanding and personalising smart city services using machine learning, the Internet-of-Things and Big Data," in *2017 IEEE 26th International Symposium on Industrial Electronics (ISIE)*, pp. 2050–2055, Edinburgh, UK, 2017.
- [7] K. Lee, D. Palsetia, R. Narayanan, M. M. A. Patwary, A. Agrawal, and A. Choudhary, "Twitter trending topic classification," in *2011 IEEE 11th International Conference on Data Mining Workshops*, pp. 251–258, Vancouver, BC, Canada, 2011.
- [8] B. O'Connor, R. Balasubramanyan, B. R. Routledge, and N. A. Smith, "From tweets to polls: linking text sentiment to public opinion time series," in *ICWSM 2010 - Proceedings of the 4th International AAAI Conference on Weblogs and Social Media*, pp. 122–129, Washington, DC, USA, 2010.
- [9] A. I. Schein, A. Popescul, L. H. Ungar, and D. M. Pennock, "DNA extraction from plant leaves with Minilys," in *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '02*, pp. 253–260, Tampere, Finland, 2002.
- [10] C. Shekar, S. Wakade, K. J. Liszka, and C. C. Chan, "Mining pharmaceutical spam from Twitter," in *2010 10th International Conference on Intelligent Systems Design and Applications*, pp. 813–817, Cairo, Egypt, 2010.
- [11] A. Zubiaga, D. Spina, V. Fresno, and R. Martínez, "Classifying trending topics: a typology of conversation triggers on Twitter," in *Proceedings of the 20th ACM international conference on Information and knowledge management*, pp. 2461–2464, Glasgow, Scotland, UK, 2011.
- [12] X. Peng, J. Feng, S. Xiao, W.-Y. Yau, J. T. Zhou, and S. Yang, "Structured autoencoders for subspace clustering," *IEEE Transactions on Image Processing*, vol. 27, no. 10, pp. 5076–5086, 2018.
- [13] X. Peng, Z. Yu, Z. Yi, and H. Tang, "Constructing the L2-graph for robust subspace learning and subspace clustering," *IEEE Transactions on Cybernetics*, vol. 47, no. 4, pp. 1053–1066, 2017.
- [14] F. Karim, S. Majumdar, H. Darabi, and S. Chen, "LSTM fully convolutional networks for time series classification," *IEEE Access*, vol. 6, pp. 1662–1669, 2017.
- [15] D. A. Shamma, L. Kennedy, and E. F. Churchill, "Peaks and persistence: modeling the shape of microblog conversations," in *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*, pp. 355–358, Hangzhou, China, 2011.
- [16] J. Yang and J. Leskovec, "Patterns of temporal variation in online media," in *Proceedings of the 4th ACM International Conference on Web Search and Data Mining, WSDM 2011*, pp. 177–186, Hong Kong, China, 2011.
- [17] A. Conneau, H. Schwenk, L. Barrault, and Y. Lecun, "Very deep convolutional networks for text classification," in *15th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2017 - Proceedings of Conference*, vol. 1, pp. 1107–1116, Valencia, Spain, 2016.
- [18] X. Zhang, J. Zhao, and Y. Lecun, "Character-level convolutional networks for text classification," *Advances in neural information processing systems*, vol. 28, pp. 649–657, 2015.
- [19] Y. Kim, "Convolutional neural networks for sentence classification," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1746–1751, Doha, Qatar, 2014.
- [20] Y. Zhang and B. Wallace, "A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification," in *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Taipei, Taiwan, 2015.
- [21] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," in *1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings*, pp. 1–12, Scottsdale, AZ, USA, 2013.
- [22] 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)*, pp. 1532–1543, Doha, Qatar, 2014.
- [23] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching word vectors with subword information," *Transactions of the Association for Computational Linguistics*, vol. 5, pp. 135–146, 2017.
- [24] D. Zeng, K. Liu, S. Lai, G. Zhou, and J. Zhao, "Relation classification via convolutional deep neural network," in *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, Dublin, Ireland, 2014.
- [25] A. Visvizi, M. D. Lytras, E. Damiani, and H. Mathkour, "Policy making for smart cities: innovation and social inclusive economic growth for sustainability," *Journal of Science and Technology Policy Management*, vol. 9, no. 2, pp. 126–133, 2018.
- [26] E. P. Trindade, M. P. F. Hinnig, E. M. da Costa, J. S. Marques, R. C. Bastos, and T. Yigitcanlar, "Sustainable development of smart cities: a systematic review of the literature," *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 3, no. 1, 2017.
- [27] N. Komninos, "Intelligent cities: towards interactive and global innovation environments," *International Journal of Innovation and regional development*, vol. 1, no. 4, p. 337, 2009.
- [28] M. Lytras, N. R. Aljohani, A. Hussain, J. Luo, and J. X. Zhang, *Cognitive Computing Track Chairs' Welcome & Organization*, pp. 247–250, 2018.
- [29] H. Mora, R. Pérez-delHoyo, J. F. Paredes-Pérez, and R. A. Mollá-Sirvent, "Analysis of social networking service data for smart urban planning," *Sustainability*, vol. 10, no. 12, p. 4732, 2018.
- [30] M. D. Lytras, H. I. Mathkour, H. Abdalla, W. Al-Halabi, C. Yanez-Marquez, and S. W. M. Siqueira, "An emerging - social and emerging computing enabled philosophical paradigm for collaborative learning systems: toward high effective next generation learning systems for the knowledge society," *Computers in Human Behavior*, vol. 51, pp. 557–561, 2015.
- [31] M. D. Lytras, W. Al-Halabi, J. X. Zhang, R. A. Haraty, and M. Masud, "Enabling technologies and business infrastructures for next generation social media: Big Data, cloud computing, Internet of Things and virtual reality," *The Journal of Universal Computer Science*, vol. 21, no. 11, pp. 1379–1384, 2015.
- [32] T. Sakaki, M. Okazaki, and Y. Matsuo, "Tweet analysis for real-time event detection and earthquake reporting system development," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 4, pp. 919–931, 2013.
- [33] A. Rosi, M. Mamei, F. Zambonelli, S. Dobson, G. Stevenson, and J. Ye, "Social sensors and pervasive services: approaches and perspectives," in *2011 IEEE International Conference on Pervasive Computing and Communications Workshops*,

- PERCOM Workshops 2011*, pp. 525–530, Seattle, WA, USA, 2011.
- [34] G. Anastasi, M. Antonelli, A. Bechini et al., “Urban and social sensing for sustainable mobility in smart cities,” in *2013 Sustainable internet and ICT for sustainability, SustainIT 2013*, Palermo, Italy, 2013.
- [35] A. Crooks, A. Croitoru, A. Stefanidis, and J. Radzikowski, “Earthquake: twitter as a distributed sensor system,” *Transactions in GIS*, vol. 17, no. 1, pp. 124–147, 2013.
- [36] E. D’Andrea, P. Ducange, B. Lazzerini, and F. Marcelloni, “Real-time detection of traffic from twitter stream analysis,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 4, pp. 2269–2283, 2015.
- [37] A. Signorini, A. M. Segre, and P. M. Polgreen, “The use of Twitter to track levels of disease activity and public concern in the U.S. during the influenza A H1N1 pandemic,” *PLoS One*, vol. 6, no. 5, article e19467, 2011.
- [38] M. Al-Ghalibi, A. Al-Azzawi, and K. Lawonn, “NLP based sentiment analysis for Twitter’s opinion mining and visualization,” in *Eleventh International Conference on Machine Vision (ICMV 2018)*, p. 6, Munich, Germany, 2019.
- [39] F. Abid, M. Alam, M. Yasir, and C. Li, “Sentiment analysis through recurrent variants latterly on convolutional neural network of Twitter,” *Future Generation Computer Systems*, vol. 95, pp. 292–308, 2019.
- [40] F. Bravo-Marquez, M. Mendoza, and B. Poblete, “Combining strengths, emotions and polarities for boosting Twitter sentiment analysis,” in *Proceedings of the second international workshop on issues of sentiment discovery and opinion mining*, pp. 1–9, Chicago, IL, USA, 2013.
- [41] T. Finin, W. Murnane, A. Karandikar, N. Keller, J. Martineau, and M. Dredze, “Annotating named entities in Twitter data with crowdsourcing,” in *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon’s Mechanical Turk*, vol. 2010, pp. 80–88, Los Angeles, CA, USA, 2010.
- [42] S. Collovini, B. Pereira, H. D. P. dos Santos, and R. Vieira, “Annotating relations between named entities with crowdsourcing,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10859 LNCS, pp. 290–297, 2018.
- [43] K. Roberts, M. A. Roach, J. Johnson, J. Guthrie, and S. M. Harabagiu, “EmpaTweet: annotating and detecting emotions on twitter,” in *Proceedings of the 8th International Conference on Language Resources and Evaluation, LREC 2012*, pp. 3806–3813, Istanbul, Turkey, 2012.
- [44] L. Barbosa and J. Feng, “Robust sentiment detection on twitter from biased and noisy data,” *Coling 2010: Posters*, vol. 2, pp. 36–44, 2010.
- [45] M. Avvenuti, S. Cresci, M. N. La Polla, A. Marchetti, and M. Tesconi, “Earthquake emergency management by social sensing,” in *2014 IEEE International Conference on Pervasive Computing and Communication Workshops, PERCOM WORKSHOPS 2014*, pp. 587–592, Budapest, Hungary, 2014.
- [46] S. M. Weiss, N. Indurkhyia, T. Zhang, and F. J. Damerau, *Text Mining: Predictive Methods for Analyzing Unstructured Information*, Springer New York, 2005.
- [47] R. Zafarani, M. A. Abbasi, and H. Liu, *Social Media Mining: An Introduction*, vol. 9781107018, Cambridge University Press, Cambridge, 2014.
- [48] B. Ait Hammou, A. Ait Lahcen, and S. Mouline, “Towards a real-time processing framework based on improved distributed recurrent neural network variants with fastText for social big data analytics,” *Information Processing & Management*, vol. 57, no. 1, article 102122, 2020.
- [49] A. S. M. Alharbi and E. de Doncker, “Twitter sentiment analysis with a deep neural network: an enhanced approach using user behavioral information,” *Cognitive Systems Research*, vol. 54, pp. 50–61, 2019.
- [50] L. M. Ang, K. P. Seng, A. M. Zungeru, and G. K. Ijemeru, “Big sensor data systems for smart cities,” *IEEE Internet of Things Journal*, vol. 4, no. 5, pp. 1259–1271, 2017.
- [51] L. M. Ang and K. P. Seng, “Big sensor data applications in urban environments,” *Big Data Research*, vol. 4, pp. 1–12, 2016.
- [52] Y. Zheng, L. Capra, O. Wolfson, and H. Yang, “Urban computing: concepts, methodologies, and applications,” *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 5, no. 3, 2014.
- [53] K. S. Tai, R. Socher, and C. D. Manning, “Improved semantic representations from tree-structured long short-term memory networks,” in *ACL-IJCNLP 2015 - 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, Proceedings of the Conference*, vol. 1, pp. 1556–1566, Beijing, China, 2015.
- [54] B. Liu, “Sentiment analysis and subjectivity,” *Handbook of Natural Language Processing*, 627–666, 2010.
- [55] A. M. Popescu and O. Etzioni, “Extracting product features and opinions from reviews,” in *HLT/EMNLP 2005 - Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, pp. 339–346, Vancouver, British Columbia, Canada, 2005.
- [56] R. Collobert and J. Weston, “A unified architecture for natural language processing,” in *Proceedings of the 25th international conference on Machine learning - ICML ’08*, pp. 160–167, Helsinki, Finland, 2008.
- [57] L. Flekova, O. Ferschke, and I. Gurevych, “UKPDIPF: lexical semantic approach to sentiment polarity prediction in Twitter data,” in *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pp. 704–710, Dublin, Ireland, 2014.
- [58] J. Serrano-Guerrero, J. A. Olivas, F. P. Romero, and E. Herrera-Viedma, “Sentiment analysis: a review and comparative analysis of web services,” *Information Sciences*, vol. 311, pp. 18–38, 2015.
- [59] C. N. Dos Santos and M. Gatti, “Deep convolutional neural networks for sentiment analysis of short texts,” in *COLING 2014 - 25th International Conference on Computational Linguistics, Proceedings of COLING 2014: Technical Papers*, pp. 69–78, Dublin, Ireland, 2014.
- [60] F. Yu and V. Koltun, *Multi-Scale Context Aggregation by Dilated Convolutions*, 2015.
- [61] N. Kalchbrenner, L. Espeholt, K. Simonyan, A. van den Oord, A. Graves, and K. Kavukcuoglu, *Neural Machine Translation in Linear Time*, 2016.
- [62] A. V. Oord, S. Dieleman, H. Zen et al., *WaveNet: A Generative Model for Raw Audio*, 2016.
- [63] R. Jozefowicz, O. Vinyals, M. Schuster, N. Shazeer, and Y. Wu, *Exploring the Limits of Language Modeling*, 2016.

- [64] M. E. Peters, W. Ammar, C. Bhagavatula, and R. Power, “Semi-supervised sequence tagging with bidirectional language models,” in *ACL 2017 - 55th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)*, vol. 1, pp. 1756–1765, Sanya, China, 2017.
- [65] M. E. Peters, M. Neumann, M. Iyyer et al., “Deep contextualized word representations,” in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pp. 2227–2237, New Orleans, LA, USA, 2018.
- [66] W. Yin and H. Schütze, *Multichannel Variable-Size Convolution for Sentence Classification*, 2016.
- [67] T. H. Nguyen and R. Grishman, “Event detection and domain adaptation with convolutional neural networks,” in *ACL-IJCNLP 2015 - 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, Proceedings of the Conference*, vol. 2, pp. 365–371, Beijing, China, 2015.
- [68] Y. Chen, L. Xu, K. Liu, D. Zeng, and J. Zhao, *Event Extraction via Dynamic Multi-Pooling Convolutional Neural Networks*, 2015.
- [69] W. Lei, K. Khine, N. Thwet, and T. Aung, *Sentiment Aware Word Embedding Approach for Sentiment Analysis*, 2018.
- [70] J. Acosta, N. Lamaute, M. Luo, E. Finkelstein, and A. Cotoranu, “Sentiment analysis of Twitter messages using Word2Vec,” in *Proceedings of Student-Faculty Research Day, CSIS, Pace University*, pp. 1–7, Pleasantville, NY, USA, 2017.
- [71] R. Petrolito and F. Dell’orletta, *Word Embeddings in Sentiment Analysis*, 2018.
- [72] A. Akbik, D. Blythe, and R. Vollgraf, *Contextual String Embeddings for Sequence Labeling*, 2018.
- [73] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, “BERT: pre-training of deep bidirectional transformers for language understanding,” in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 4171–4186, Minneapolis, MN, USA, 2019.
- [74] E. H. Alkhammash, J. Jussila, M. D. Lytras, and A. Visvizi, “Annotation of smart cities twitter micro-contents for enhanced citizen’s engagement,” *IEEE Access*, vol. 7, pp. 116267–116276, 2019.
- [75] A. Zubiaga, “A longitudinal assessment of the persistence of twitter datasets,” *Journal of the Association for Information Science and Technology*, vol. 69, no. 8, pp. 974–984, 2018.
- [76] J. Gu, Z. Wang, J. Kuen et al., “Recent advances in convolutional neural networks,” *Pattern Recognition*, vol. 77, pp. 354–377, 2018.
- [77] C. Gan, L. Wang, Z. Zhang, and Z. Wang, “Sparse attention based separable dilated convolutional neural network for targeted sentiment analysis,” *Knowledge-Based Systems*, vol. 188, 2019.
- [78] *Twitter event datasets (2012-2016)*, November 2019, https://figshare.com/articles/Twitter_event_datasets_2012-2016_/5100460.
- [79] TAGS Twitter Archiving Google Sheet, November 2019, <https://tags.hawksey.info/>.
- [80] C. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. Bethard, and D. McClosky, *The Stanford CoreNLP Natural Language Processing Toolkit*, pp. 55–60, 2015.
- [81] B. Chiu, G. Crichton, A. Korhonen, and S. Pyysalo, “How to train good word embeddings for biomedical NLP,” in *Proceedings of the 15th Workshop on Biomedical Natural Language Processing*, pp. 166–174, Berlin, Germany, 2016.
- [82] A. C. Ian Goodfellow and Y. Bengio, “Deep learning,” *Genetic Programming and Evolvable Machines*, vol. 19, no. 1–2, pp. 305–307, 2017.
- [83] B. Xu, N. Wang, T. Chen, and M. Li, *Empirical Evaluation of Rectified Activations in Convolutional Network*, 2015.
- [84] J. Heaton, “The number of hidden layers,” in *Introduction to Neural Networks for Java*, pp. 157–158, Heaton Research, Inc., 2008.
- [85] *TensorFlow*, 2016, November 2019, <https://www.tensorflow.org/>.