Tech Science Press

check for updates

Multi-label Emotion Classification of COVID–19 Tweets with Deep Learning and Topic Modelling

K. Anuratha[1,*] and M. Parvathy[2]

[1]Department of Information Technology, Sri Sai Ram Institute of Technology, Chennai, Tamilnadu, India
[2]Department of Computer Science and Engineering, Sethu Institute of Technology, Madurai, Tamilnadu, India
*Corresponding Author: K. Anuratha. Email: anuvinaya2003@gmail.com

Abstract: The COVID-19 pandemic has become one of the severe diseases in recent years. As it majorly affects the common livelihood of people across the universe, it is essential for administrators and healthcare professionals to be aware of the views of the community so as to monitor the severity of the spread of the outbreak. The public opinions are been shared enormously in microblogging media like twitter and is considered as one of the popular sources to collect public opinions in any topic like politics, sports, entertainment etc., This work presents a combination of Intensity Based Emotion Classification Convolution Neural Network (IBEC-CNN) model and Non-negative Matrix Factorization (NMF) for detecting and analyzing the different topics discussed in the COVID-19 tweets as well the intensity of the emotional content of those tweets. The topics were identified using NMF and the emotions are classified using pretrained IBEC-CNN, based on predefined intensity scores. The research aimed at identifying the emotions in the Indian tweets related to COVID-19 and producing a list of topics discussed by the users during the COVID-19 pandemic. Using the Twitter Application Programming Interface (Twitter API), huge numbers of COVID-19 tweets are retrieved during January and July 2020. The extracted tweets are analyzed for emotions fear, joy, sadness and trust with proposed Intensity Based Emotion Classification Convolution Neural Network (IBEC-CNN) model which is pretrained. The classified tweets are given an intensity score varies from 1 to 3, with 1 being low intensity for the emotion, 2 being the moderate and 3 being the high intensity. To identify the topics in the tweets and the themes of those topics, Non-negative Matrix Factorization (NMF) has been employed. Analysis of emotions of COVID-19 tweets has identified, that the count of positive tweets is more than that of count of negative tweets during the period considered and the negative tweets related to COVID-19 is less than 5%. Also, more than 75% negative tweets expressed sadness, fear are of low intensity. A qualitative analysis has also been conducted and the topics detected are grouped into themes such as economic impacts, case reports, treatments, entertainment and vaccination. The results of analysis show that the issues related to the pandemic are expressed different emotions in twitter which helps in interpreting the public insights during the pandemic and these results are beneficial for planning the dissemination of factual health statistics to build the trust of the people. The performance comparison

shows that the proposed IBEC-CNN model outperforms the conventional models and achieved 83.71% accuracy. The % of COVID-19 tweets that discussed the different topics vary from 7.45% to 26.43% on topics economy, Statistics on cases, Government/Politics, Entertainment, Lockdown, Treatments and Virtual Events. The least number of tweets discussed on politics/government on the other hand the tweets discussed most about treatments.

1 Introduction

Social media platforms such as Face book, Twitter have a significant role in propagation of information, expression of emotions and production of knowledge from various sources. The pandemic COVID-19 was detected in 2019 and millions of people across the globe have been affected by the virus and also caused several deaths. The pandemic not only caused illness but also affected the lives of people in various nations. Though the countries started implementing the methods to contain the disease, it is evident that different opinions were propagated by the media and press as well in the social medium. There has been a habitual stream of information in television media channels as well the social media about the virus and its variants, from the late 2019. Though the virus originated from China, in no time it spread over several countries. The clinicians got more facts about the virus and the people vulnerable to it, on observing the pattern of the spread of the outbreak. Already, there are many studies have been done on observing the people views on social media throughout the beginning stages of the crisis. Reference [1] Shows that the analysis of twitter helped health professionals in South Korea in drafting the decisions on heath policies. The social media analysis, in particular the microblogging site twitter analysis provide significant information on the trending news in real time like pandemic, sports, politics etc., as well the public opinions on that news. In the year 2020, the COVID-19 pandemic became a topic that trends daily in the twitter. Many people right from a commoner to the first citizen of the nation started tweeting on COVID-19, to express the different emotions related to the pandemic. The regular analysis of social platforms and the pandemic has discovered the approaches for assessing the public views, using social media platforms for policy making and for the accurate prediction of the information [2].

Natural Language Processing (NLP) and its methods have gained importance to assess the huge amount of data produced in the natural language. NLP is a domain that combines machine intelligence with linguistics, to empower machines to interpret human language. As vast amount of data of various structure is been generated in social media, NLP has become a challenging task these days. Most common NLP methods are topic modelling, opinion mining and text summarization. In the COVID-19 epidemics, it is essential to have data assessment methods to interpret the information on various topics discussed and the emotions expressed on those topics. These topics include crisis management, politics, treatment, economic impact, sports, entertainment, news etc.

Emotion analysis is analyzing the user tweets to understand the psychological state of the users, revealing the various emotions such as sadness, happy, fear, anger, trust etc., The emotions are automatically classified using a Convolutional Neural Network (CNN). Manual processing of the tweets is impossible to determine the topic on public opinions. A useful tool that automatically discovers the topic about a document from massive texts is Topic modelling which is an unsupervised machine learning technique. Earlier, the topic modelling was achieved by most common techniques such as Latent Semantic Analysis (LSA) [3], Probabilistic Latent Semantic Analysis (PLSA) [4], Latent Dirichlet

Allocation (LDA) [5] and Nonnegative Matrix Factorization (NMF) [6]. The techniques LDA, LSA and PLSA employ the assumption that the document is a collection of topics and every topic is a collection of words. The matrix document-word is split in to document–topic matrix and topic-words matrix by those techniques. The document-topic distribution illustrates each documents' degree, which belong to a specific topic and the topic-word specifies the degree of each word, which constitutes a topic. Alternatively, NMF directly decomposes the document in to two low-level factor matrices and reveal the structure of the topics that constitute a document. Conventional methods have gained success in topic modelling but suffer from performance degradation when comes to short texts like tweets. Data sparsity is also a problem with these methods. Whereas, the NMF model shows better performance for data with high-dimensional space using dimensionality reduction and clustering.

The research objectives of the work are as follows:

- To develop an approach for automatically detecting and classifying the COVID-19 tweets exploring the supervised Deep learning model and unsupervised Topic Modelling.
- To propose a CNN model IBEC-CNN to identify the COVID-19 tweets with multi-labelled emotions and classify them based on intensity level expressed in emotions.
- To discover the unique topics and issues explored in the COVID-19 tweets.
- To analyze the topics with positive emotions trust, joy and negative emotions fear, sadness.

The rest of the sections of the paper are structured as follows: Section 2 details about work focused on topic modelling and emotion classification using deep learning. Section 3 describes the proposed methodology, details on implementation. Section 4 discusses the results and performance comparison. Section 5 concludes the work with future enhancements.

2 Literature Survey

Several researchers have focused on Twitter to interpret the impact of the epidemic in civic behavior. Reference [7] shows the main topics that are discussed during the pandemic applying topic modeling technique Latent Dirichlet Allocation (LDA) and opinion mining. The individual's social survival perceptions are studied using the regular use of social platforms during the epidemic; also it led to data overload [8]. To assess the appropriateness of the social surveys on different topics, the social platforms are analyzed [9]. Few studies focused on analyzing the tweets to know the impact of the pandemic in human lives [10,11]. Recent times, People express their opinions on diverse topics over emotions. These emotions could be used for business and data analysis [12–15]. The opinions may be of positive or negative polarity and may express emotions like trust, sadness, happy and anger [16]. Though several studies focused on binary classification of opinions it's a challenging issue to classify a tweet with many emotions' terms. Owing to the epidemic and quarantine policies, during the pandemic lock down social platforms have become the main mediums for sharing the thoughts on effect of the COVID-19 in day-to-day routine. The hidden topics in the user tweets reflect the issues of the people over time, which demands temporal methods such as NMF to discover important topics such as economy impact, treatment, testing, and politics [17]. Emotion classification is the process of analyzing and classifying the user tweets, emoticons in to a group of sentiments [18]. The emotion classification can be carried out at word, sentence and document level using different supervised, semi supervised and unsupervised classification algorithms [19–21]. Twitter opinion mining is a promising application of Natural Language Processing that employs various techniques to classify the emotions in a tweet [22] and used in several use cases like healthcare, finance, entertainment, marketing etc., show the emotion classification of hate speech in tweets and discussed the topics discovered in those tweets in the Arabic region. The tweets with multiple labels of emotions with various intensity levels can be classified using the Convolutional

Neural Networks. The training efficiency and the performance of the CNN model can be improved by reducing the size of parameters and feature space [23]. Reference [24] shows how Convolutional Neural Network combined with extreme learning machine employed to classify the Electrocardiogram (ECG) signals.

The existing studies focused on the emotion classification or topic identification but only few researches have focused on intensity based emotion classification and the topic modeling was a quite difficult in terms of short texts in particular for tweets. The topic modeling can be achieved with NMF for short tweets and the same has been employed in the proposed model IBEC-CNN.

3 Proposed Methodology

The main focus of this work was on emotion classification of tweets with multiple labels and identification of tweets in those topics. The conventional models used for topic modelling suffer from data sparsity problem and performance degradation in the case of short texts such as tweets. Here, the proposed methodology used NNMF model that outperforms the other existing models for high-dimensional data applying clustering and dimensionality reduction. Also, the issue of diverse labels exist in the emotions of various classes of tweets are classified with proposed IBEC-CNN model.

The overall working model of the proposed IBEC-CNN is shown in Fig. 1. In this work, we aimed at automatically detecting the COVID-19 tweets posted from India using Convolutional Neural Network and NMF topic modeling. Initially we do the tweets extraction and preprocessing. The next phase will be classification to identify the intensities with different classes of sentiments exist in the tweets using CNN. Subsequently, the topics discussed in the classified tweets are identified using NMF Topic Modeling.
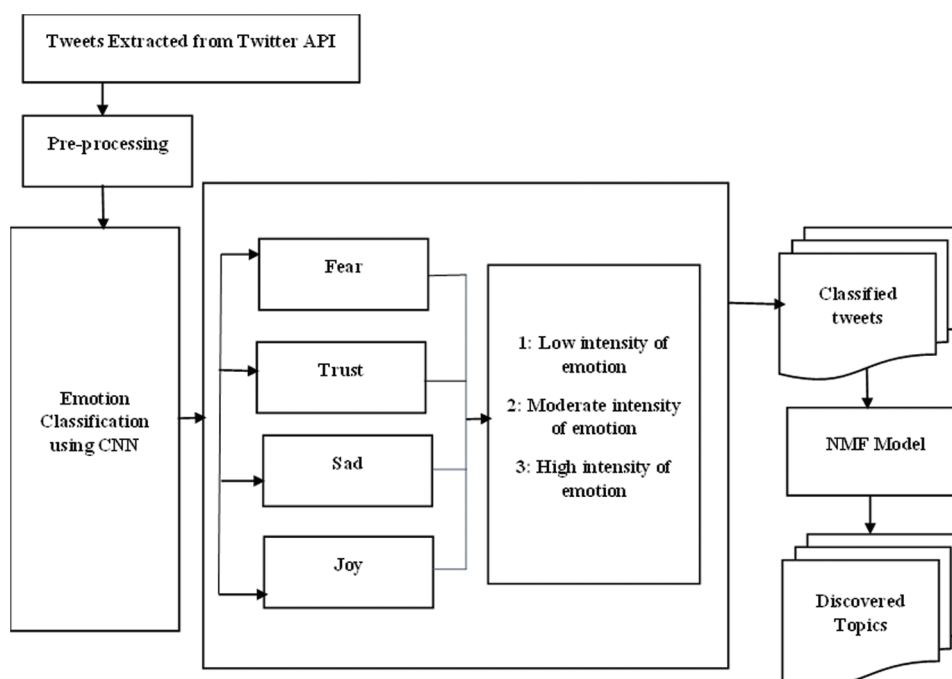

Figure 1: Block diagram–proposed methodology

3.1 Data Collection and Dataset

The COVID-19 tweets are collected by querying the Twitter API using the key words "Covid", "#pandemic", "Corona virus" etc., massive number of tweets are extracted for the period January 2020 to

July 2020. Duplicate tweets (retweets, quoted tweets) are discarded and original English tweets alone are taken for the dataset. The data corpus is built with the COVID-19 tweets extracted from the twitter using the Twitter API. The corpus has all the three classes of tweets with emotions of various intensity levels.

3.2 Preprocessing

Different people express lot of opinions on various topics over time. The twitter data has large amount of data with multiple attributes. Preprocessing of the tweets is done to remove stop words, Uniform Resource Locator (URL), emojis, symbols, punctuations, whitespace. The emojis are replaced with the description. Hashtags are preserved as they may have topical patterns hidden.

3.3 Emotion Classification Using CNN

The model we used for emotion classification is a CNN model as the application of the model to classify texts is efficient. The proposed CNN model for emotion classification is shown in Fig. 2 has five layers: an input, two convolutions, pooling and the output layer. The 5 layers are intended for the functions: word vectorization, sentence vectorization, tweet vectorization and finally the classification.
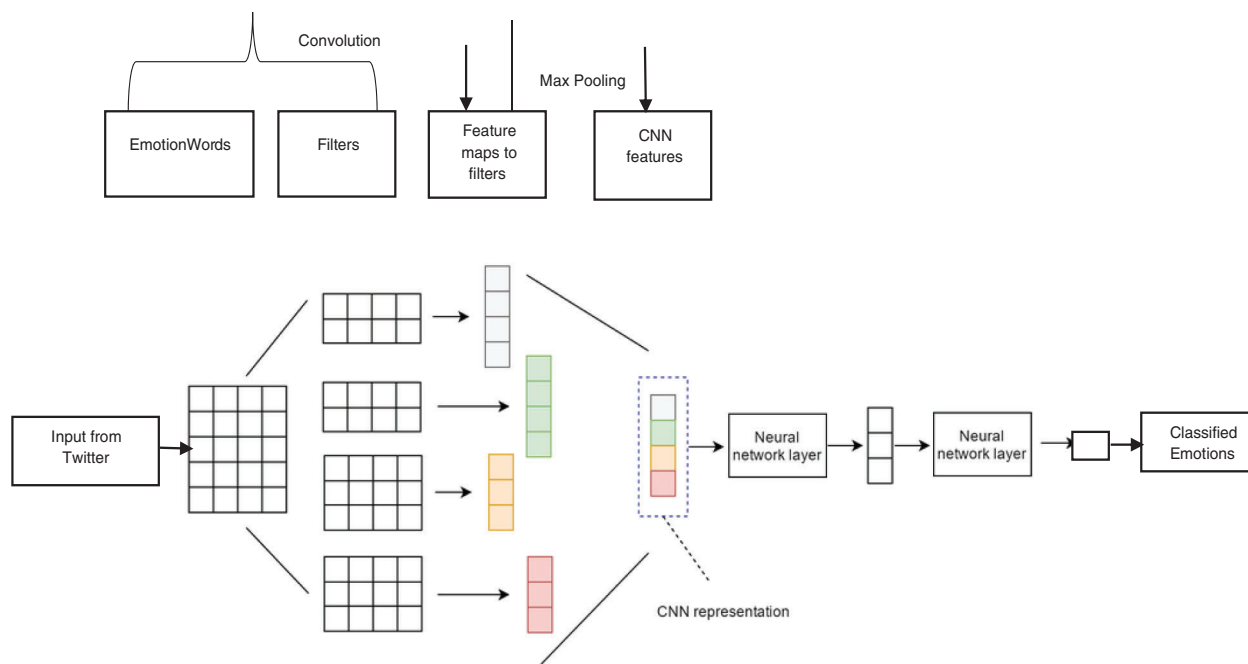

Figure 2: CNN architecture–emotion classification

Tokens are used to represent the long sentences in short form. A definite length L has been assigned to the longest tweet. The mapping of every term of a sentence to the feature space is done using the embedding layer and results in $L \times E$, where E says the size of embedding. The embedding layer holds the terms that are semantically related adjacently and the rest are positioned far away, which is typically a process to attain 2D-word vector matrix. These convolution features are then fed in to max pooling layer to slide the vectors using a random slide every value of a matrix that is the result of subsequent convolution layer, using a random pace. The convolution layer is responsible for selecting the significant features from the input vector to produce the feature space. The transformation of 2D feature space to 1D is done with flattening of the resultant feature

matrix. This 1D feature space is the input to the fully connected layer and the fully connected layer associates each input to the output neurons.

The fully connected layer creates a set of positive and negative emotions categorized using a vector and with the SoftMax activation function. The model has three CNNs, convolution layer and max-pooling layers and three of them apply the same activation function as well the same number of filters on the input tweets but the convolution layers and the max-pooling layers are diverse in size to one another. The Co1 × E is the initial convolution, which has the terms to undergo emotion classification in l × E and the passes the data to subsequent convolution layer. The next convolution layer uses Co2 × E for the extraction of significant features from the resultant data of Co1 × E. Co1 and Co2 are the sizes of the filters used in each convolution layer and are restricted to C1, C2 different filters respectively. The first convolution layer is applied to search for contextual words in L × E and the next convolution layer is used for extracting the essential features discarding the irrelevant features that may affect the classification accuracy. The output layer associates each text with a level varies from 1 to 3 with 1 being the low intensity of the emotion expressed and 3 being the high intensity. We have used threshold 50% or 0.5 to classify an emotion with a value 0.5 or more as positive and any probability less than 0.5 as negative. The emotions are further divided in to three intensities low (0.5 to 0.65), moderate (0.66 to 0.80) and high (0.81 to 1.00) based on the threshold of the detected tweets.

3.4 Non-Negative Matrix Factorization Model for Topic Discovery

There are two approaches for topic modelling: LDA and NMF [6]. The Non-negative Matrix Factorization (NMF) approach is shown in Fig. 3 and is best suitable for extracting the topics from the tweets [25] and the same has been used in our work. The NMF approach is an unsupervised technique which learns the high-dimensional data and transforms in to representations of low-dimensions. This is the way it reduces the dimensionality of non-negative vectors. Its potential to detect the latent relationships of texts and discover the hidden topics within the emotions expressed in the texts attract researchers in the field of emotion classification. In this work, we employed the NMF approach to understand the semantics of the textual structure and discover the latent topics within the emotions as its ability to discover more clear topics than the conventional technique for topic discovery, LDA approach. NMF gives two matrices W and H from the original matrix V. W represents the topics found and H illustrates the associated weights for the topics. H is documents by topics and W is topics by words.
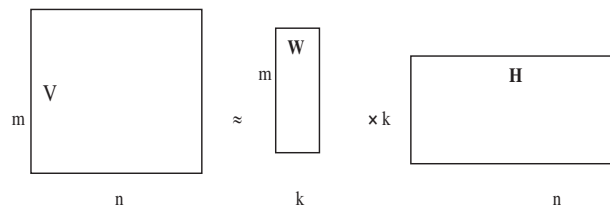

Figure 3: NMF approach

$$W \times H \approx V \tag{1}$$

The corpus of tweets is considered as a matrix of rows and columns where rows represent topics and columns represent associated weights. The topic word distribution is exploited using $W \in k \times n$ as the hidden topics are represented in columns and the document topic distribution is exploited using $H \in k \times n$ as each column indicates topic weight distribution of each tweet. The entire tweet corpus is divided into several small batches and eventually used for recoding W and H of each batch.

The matrix W is of size m × k and H is of size k × n and W × H ≈ V. The decomposition of V in to W and H is not distinct as the value k provides varied W and H. The NMF technique is applied to the preprocessed tweets, transforming the preprocessed and classified features into unigrams and bigrams (TF-IDF vectors). In this study, the tweet corpus is constructed with words and tweets represented as rows and columns respectively.

3.5 Discovering Topics on Emotions

The extracted tweets are analysed referring the timestamp, unique ID and topics in which different emotions expressed to monitor the changes in emotions, topics over time to be aware of the perceptions of public in the COVID-19 epidemic.

4 Results and Discussions

The initial process is to analyze the extracted tweets to detect the emotions using the Intensity Based Emotions-Convolutional Neural Network Model. The total number of COVID-19 tweets detected with positive and negative emotions is presented in Tab. 1. It shows the intensity of emotions expressed in the tweets as well. It is shown that out of the collected COVID-19 tweets, 96.7% of tweets expressed positive emotions and only 3.3% of tweets expressed negative emotions. The intensity level of the negative tweets is low for 76.3% of the negative tweets and 2.33% of tweets are of high intensity.

Table 1: Total number of COVID-19 tweets with positive and negative emotions

Emotion class		No of tweets
Negative		21539
Intensity of emotion	Low	16434
	Moderate	4603
	High	502
Positive		631166
Intensity of emotion	Low	284249
	Moderate	189350
	High	157567
Total No. of tweets		652705

4.1 Changes in Emotions Over Time

The COVID-19 tweets posted during the period January and July 2020 are collected for the experimental analysis. We have divided these tweets in to three durations: January to February, March to May, June to July. Tab. 2 gives the distribution of COVID-19 tweets over the three durations with the positive and negative class along with the intensity levels. It is observed that during the first time period January to March there was around 95.2% tweets discusses about COVID-19 and there was a sudden increase by about 104.2% during the second duration. There was a fall by about 29.5% during mid of second duration and 33.1% in the third duration. The negative emotions expressed in the COVID-19 tweets from the mid of first duration to the mid of second duration increased by 103.3% and from the second half to third duration it decreased by 53.8%. However, its observed that the number of COVID-19 cases kept on increasing over the three durations taken for consideration. Looking at the intensity of those extracted tweets, the

intensity of positive emotions decreased over the period of March to May and increased over June to July but the intensity Moderate and High with a minimum % of increase. On observing the statistics, it is evident that the number of negative tweets increases as there is increase in the COVID-19 cases and the positive tweets decrease comparatively. The intensity level is varying at different rates across the three durations for both the class of emotions. From the intensity levels it is clear that the number of negative tweets and the rate of spread of COVID-19 are weakly associated over time. The statistics on number of COVID-19 tweets and the classes of emotions expressed in those tweets over different time duration is illustrated in Fig. 4.

Table 2: Total number of COVID-19 tweets with positive and negative emotions over different time duration

| Duration		Jan–Feb	March–May	June–July
COVID-19 tweet categories		286116	155723	210866
Negative		3815	11416	6308
Intensity	Low	2493	9405	4536
	Moderate	1157	1808	1638
	High	165	203	134
Average % of Tweets with negative emotions		1.33	7.33	2.99
Positive		282301	144307	204558
Intensity	Low	111968	62950	107331
	Moderate	107992	32642	48716
	High	62341	48715	48511
Average % of Tweets with positive emotions		98.67	92.67	97.01
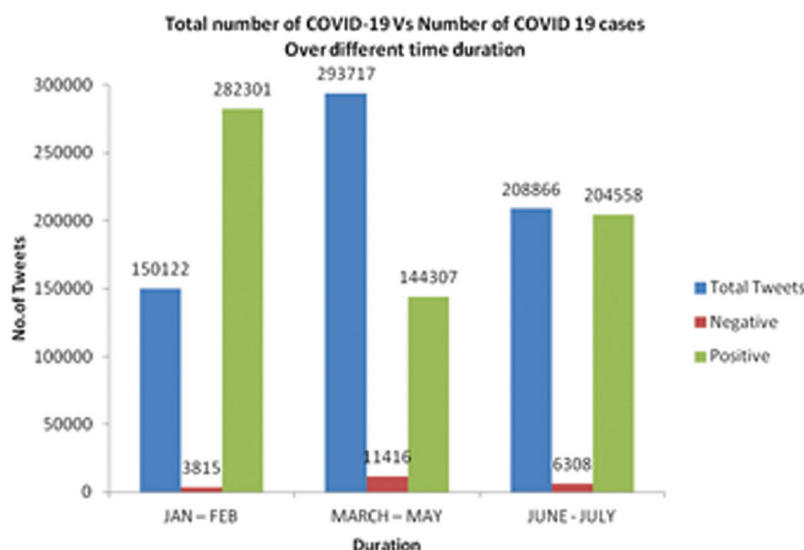
Figure 4: Total number of COVID-19 *vs.* number of COVID 19 cases over different time duration

Tab. 3 shows the measures on performance for the affective dimension. It is observed that the emotion joy is predicted with the maximum accuracy of 89.32%, emotion sadness with 87.21%, emotion fear with 81.65% and emotion trust with 79.34%. The precision measure is maximum for sadness emotion with

94.31%, joy with 75.16%, trust with 67.31% and fear with 62.18%. On the average the IBEC-CNN model achieved 84.38% of accuracy, 67.41% of F1, 64.28% of recall and 74.74% of precision. Fig. 5 illustrates the % of performance achieved by the proposed model for the different measures.

Table 3: Performance measure using the affective dimension–proposed IBEC-CNN model

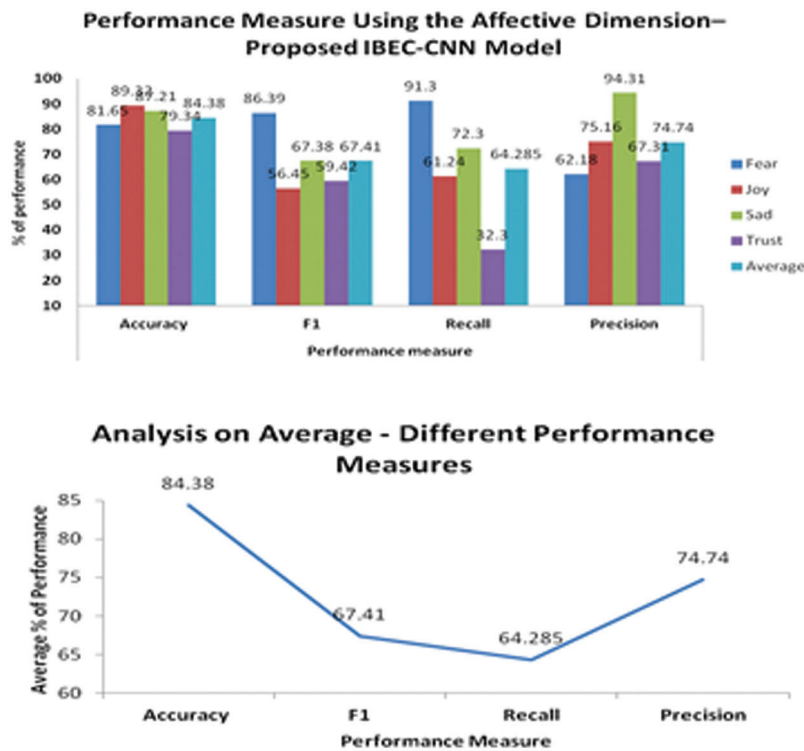|		Performance measure			
		Accuracy	F1	Recall	Precision
Emotion	Fear	81.65	86.39	91.3	62.18
	Joy	89.32	56.45	61.24	75.16
	Sadness	87.21	67.38	72.3	94.31
	Trust	79.34	59.42	32.3	67.31
Average		**84.38**	**67.41**	**64.285**	**74.74**

Figure 5: Performance measure using the affective dimension–proposed IBEC-CNN model

The following Tab. 4 shows the analysis on classification achieved by the IBEC-CNN model for different intensity levels of positive and negative classes. The classification performance under the measures accuracy, precision, recall and F1 under for the emotion classes fear, joy, sadness and trust is illustrated in Fig. 6.

Table 4: Performance measure–analysis on intensity levels

			Performance measure			
			Accuracy	F1	Recall	Precision
Emotion with intensity	Fear	Low	87.81	48.65	35.28	91.62
		Moderate	86.21	71.02	67.25	78.34
		High	89.67	75.21	46.21	86.31
	Average		**87.89**	**64.96**	**49.58**	**85.42**
	Joy	Low	83.45	64.28	59.84	68.75
		Moderate	72.56	61.65	89.87	49.27
		High	79.61	32.87	27.28	67.24
	Average		**78.54**	**54.55**	**52.32**	**63.14**
	Sadness	Low	84.25	69.41	56.91	89.71
		Moderate	82.5	76.57	84.08	71.82
		High	79.24	64.20	52.74	83.72
	Average		**81.9**	**67.4**	**56.50**	**78.14**
	Trust	Low	92	74.85	68.87	92.03
		Moderate	83.2	71.25	62.41	89.41
		High	91.45	68.52	55.92	92.84
	Average		**86.49**	**68.51**	**54.87**	**85.39**

The average performance measure analysis across the different emotions at various intensity levels and the affective dimension is shown in Tab. 5 and visualized in Fig. 7.

The comparative analysis on the proposed IBEC-CNN model with the conventional models is presented in Tab. 6 and Fig. 8. From the analysis it is inferred that the proposed model outperforms the existing classifiers with a accuracy % of 83.71.

4.2 Topics Discussed in COVID-19 Tweets

The research has considered seven topics for the analysis and the % of COVID-19 tweets discussed these topics identified by the NMF Model with the associated key terms are illustrated in Tab. 7 and Fig. 9.

The Tab. 8, Figs. 10 and 11 show the distribution of COVID-19 tweets discussed a specific topic under different intensity levels of positive and negative emotions.

Tab. 9 describes analysis of time taken by the proposed model compared with other existing models and the proposed model takes less time.

Across the years, many researches have been done on emotion classification of tweets. The process of classifying emotions is complex and with multiple intensities make it still more complex in terms of achieving better accuracy. This research proposed a CNN model integrated with NMF topic modeling, to accurately classify the short texts on COVID-19 that exhibit various emotions at diverse levels. When comparing with other models, the proposed model IBEC-CNN improved the performance measures significantly and found to be efficient for classifying multi-label sentiments.
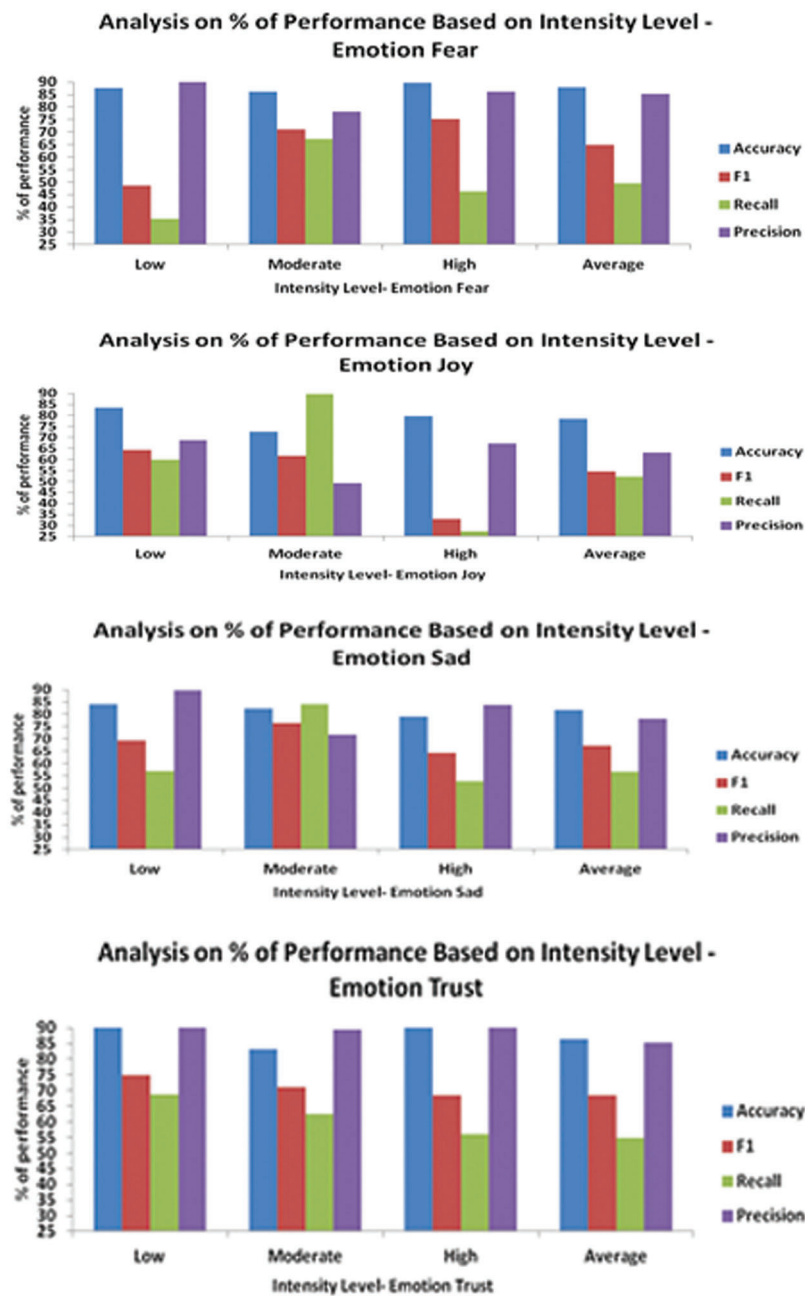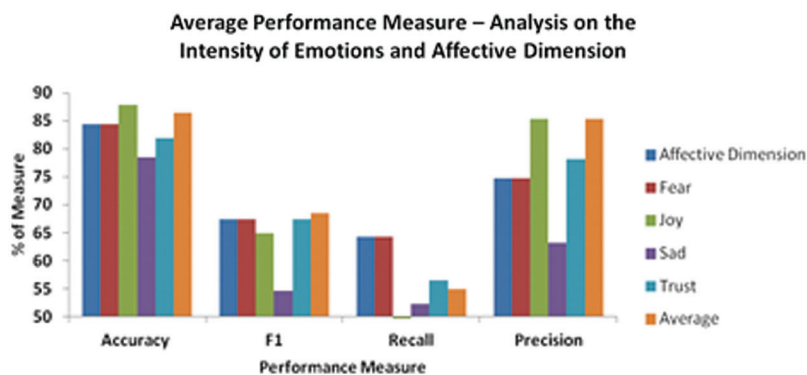
Figure 6: Performance measure–analysis of intensity levels

Table 5: Average performance measure–analysis on the intensity of emotions and affective dimension

		Performance measure			
		Accuracy	F1	Recall	Precision
Affective dimension		84.38	67.41	64.285	74.74
Emotion	Fear	87.89	64.96	49.58	85.42
	Joy	78.54	54.55	52.32	63.14
	Sadness	81.9	67.4	56.50	78.14
	Trust	86.49	68.51	54.87	85.39
Average		**83.71**	**63.86**	**53.32**	**78.02**

Figure 7: Average performance measure–analysis on the intensity of emotions and affective dimension

Table 6: Performance measure–comparison on IBEC–CNN with existing models

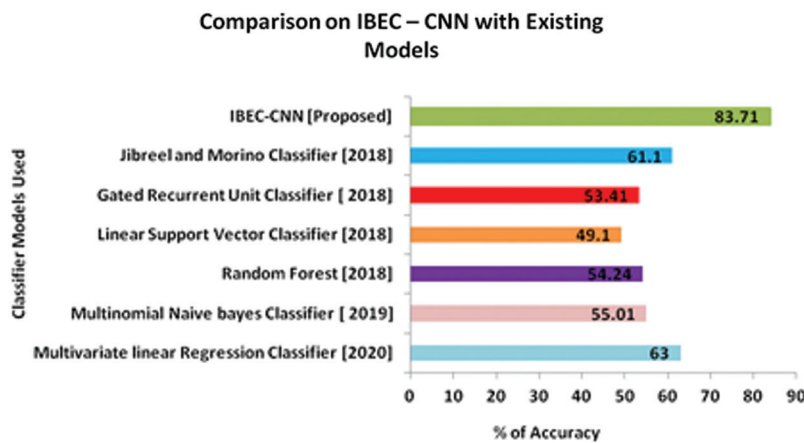| Models	Performance accuracy (in %)
Multivariate Linear Regression Classifier [2020]	63
Multinomial Naïve Bayes Classifier [2019]	55.01
Random Forest [2018]	54.24
Linear Support Vector Classifier [2018]	49.10
Gated Recurrent Unit Classifier [2018]	53.41
Jibreel and Morino Classifier [2018]	61.1
IBEC-CNN [Proposed]	**83.71**

Figure 8: Performance measure–comparison on IBEC–CNN with existing models

Table 7: NMF topics list with associated keywords in tweets

| Topic ID	Topics	Keywords associated	% of Tweets
1	Economy	Salary, recession, business, work from home, consequence, employment	8.32
2	Statistics on cases	Death, increase in patients in ICU, infection rate, isolation	12.21
3	Government/ Politics	State minister, corporation, district, safety protocols, chief minister, response to cases	7.45
4	Entertainment	Videos, Netflix, you tube, amazon prime, OTT, DTH, online games	11.06
5	Lockdown	Isolation, quarantine, containment, travel pass, regulations on travel, red zone, social distance	25.32
6	Treatments	Hospitals, ICU, ventilation, Vaccination, masks, medicines, prevention, sanitizer, handwash,	26.43
7	Virtual events	Online classes, zoom, gmeet, host, live, video	9.21
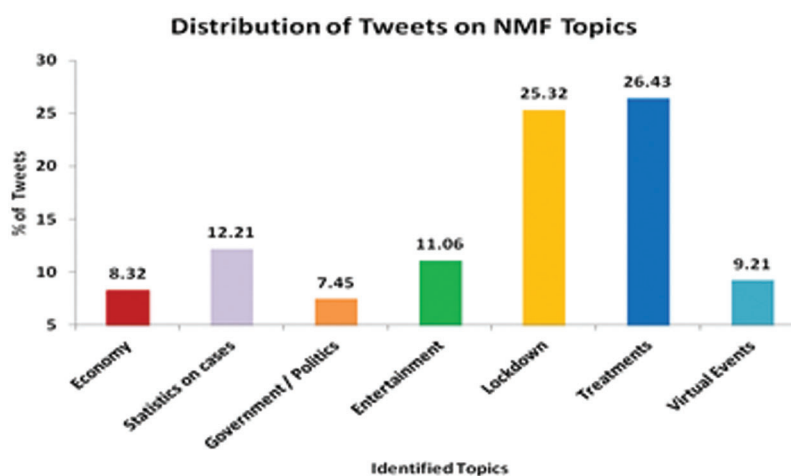
Figure 9: Distribution of tweets on NMF topics

Table 8: NMF topics list–distribution of tweets by emotions on different topics

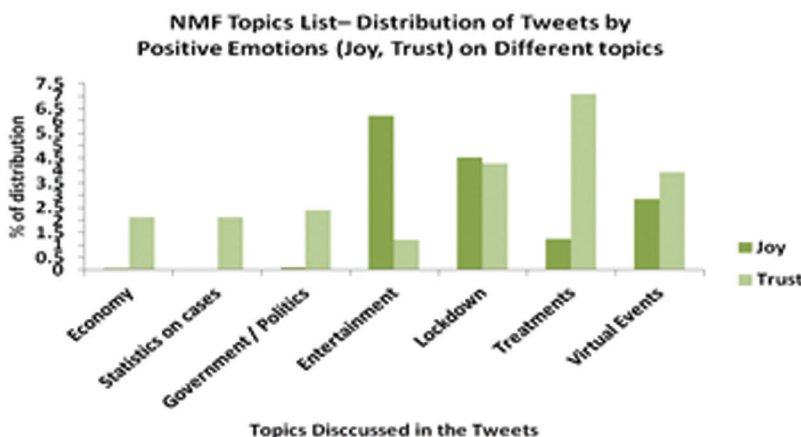| Topic ID	Topics	Emotions (% of Tweets)				Total % of Tweets on a specific topic
		Fear	Joy	Sadness	Trust	
1	Economy	2.82	0.07	3.82	2.11	8.32
2	Statistics on cases	4.81	0.01	5.27	2.13	12.21
3	Government/Politics	2.6	0.1	2.35	2.4	7.45
4	Entertainment	0.08	6.2	3.58	1.2	11.06
5	Lockdown	3.28	4.54	13.2	4.3	25.32
6	Treatments	8.32	1.24	9.8	7.07	26.43
7	Virtual Events	1.2	2.87	1.2	3.94	9.21

Figure 10: NMF topics list–distribution of tweets by positive emotions (Joy, Trust) on different topics
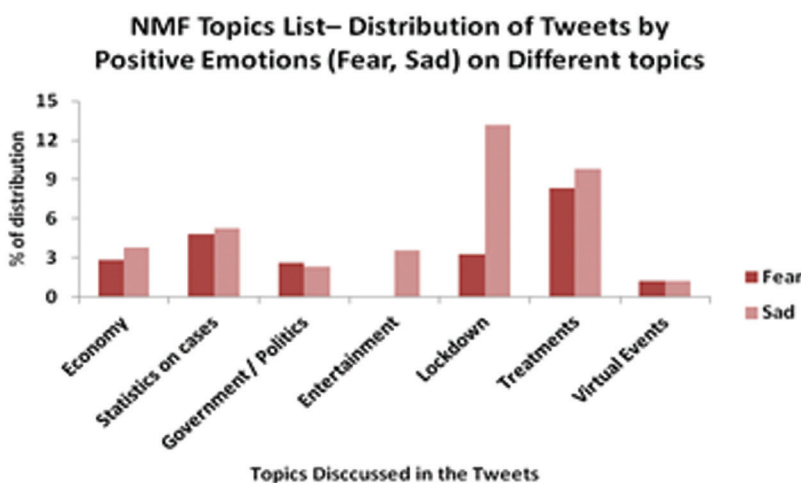
Figure 11: NMF topics list–distribution of tweets by negative emotions (Fear, Sadness) on different topics

Table 9: Analysis on time complexity

Analysis of time complexity in milliseconds (ms)	
Multivariate Linear Regression Classifier [2020]	96.11
Multinomial Naive Bayes Classifier [2019]	92.67
Random Forest [2018]	87.22
Linear Support Vector Classifier [2018]	63.21
Gated Recurrent Unit Classifier [2018]	92.43
Jibreel and Morino Classifier [2018]	88.45
IBEC-CNN [Proposed]	**58.86**

5 Conclusion and Future Work

Across the world, the COVID-19 epidemic has become the vital threat for the health sector. It also affects the general public in various aspects right from mental health to economy. This work analyzed the emotions expressed in the COVID-19 tweets extracted from the Twitter API using unsupervised technique topic modeling and supervised deep learning model. It is observed from the research that the number of positive tweets is extremely more than that of negative tweets. It is also evident that maximum negative tweets are at intensity level low. The work has considered three durations for the analysis in which the second duration (March–May) has the highest number of negative tweets which is 53% in over all negative tweets. It is also found that the tweets as well the intensity levels on the tweets are not in line with the trend in raise or decline in COVID-19 cases. This work also discovered the topics explored by the twitter users under different emotions, during the pandemic. On analyzing the results of Nonnegative Matrix Factorization topic modeling, it is understood that people have exhibited mixed emotions i.e., positive (joy, trust) and negative (sadness, fear) emotions on the identified topics. We have used the IBEC-CNN model for classifying the tweets which outperformed all the conventional classifiers with a performance accuracy of 83.71% predicting the emotions based on the intensities as well and also the topics discussed in those tweets and can be used as tool for classifying emotions in the social media posts. On understanding the different emotions on COVID-19 tweets, customized and intended facts could be developed to disseminate the reliable health information and will be useful for the administrators as well the clinicians to better administer the pandemic. The limitation of the work is that, it considered only the English tweets originated from India for specific period.

Future work could investigate diverse algorithms to analyze the regional languages. The proposed work can be enhanced to examine the COVID 19 tweets across diverse countries and to present a comparative analysis on the psychological changes, emotions and consequences of COVID-19 in various aspects.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] H. W. Park, S. Park and M. Chong, "Conversations and medical news frames on twitter: Endemiological study on COVID-19," *South Korea Journal Medical Internet Resource*, vol. 22, no. 5, pp. 88–197, 2020.

[2] P. Fortuna and S. Nunes, "A survey on automatic detection of hate speech in text," *ACM Computer Survey*, vol. 51, no. 4, pp. 1–30, 2018.

[3]   S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer and R. Harshman, "Indexing by latent semantic analysis," *Journal American Social Information Science*, vol. 41, no. 6, pp. 391–407, 1990.

[4]   T. Hofmann, "Probabilistic latent semantic indexing," *Proc. 22nd Annual Int. ACM SIGIR Conf. Resource Development*, vol. 1, pp. 50–57, 1999.

[5]   D. M. Blei, A. Y. Ng and M. I. Jordan, "Latent Dirichlet allocation," *Journal Mechanical Learning Resource*, vol. 3, pp. 993–1022, 2003.

[6]   D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Journal of Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[7]   S. J. Hoffman and V. Justicz, "Automatically quantifying the scientific quality and sensationalism of news records mentioning pandemics: Validating a maximum entropy machine-learning model," *Journal of Clinical Epidemiology*, vol. 75, no. 4, pp. 47–55, 2016.

[8]   A. Abd-Alrazaq, D. Alhuwail, M. Househ, M. Hamdi and Z. Shah, "Top concerns of tweeters during the COVID-19 pandemic: Infoveillance study," *Journal of Medical Internet Resources*, vol. 22, no. 4, pp. 116–190, 2020.

[9]   A. Farooq, S. Laato and A. Islam, "Impact of online information on self-isolation intention during the COVID-19 pandemic: A cross-sectional study," *Journal of Medical Internet Resources*, vol. 22, no. 5, pp. 190–196, 2020.

[10] W. Jo, J. Lee, J. Park and Y. Kim, "Online information exchange and anxiety spread in the early stage of the novel coronavirus (COVID-19) outbreak in South Korea: Structural topic model and network analysis," *Journal of Medical Internet Resources*, vol. 22, no. 6, pp. 116–190, 2020.

[11] H. Budhwani and R. Sun, "Creating COVID-19 stigma by referencing the novel coronavirus as the "Chinese virus" on Twitter: Quantitative analysis of social media data," *Journal of Medical Internet Resources*, vol. 22, no. 5, pp. e19301, 2020.

[12] Y. Chen, Y. Zhou, S. Zhu and H. Xu, "Detecting offensive language in social media to protect adolescent online safety," in *Proc. of the 2012 Int. Conf. on Privacy, Security, Risk and Trust (PASSAT)*, Amsterdam, vol. 1, pp. 71–80, 2012.

[13] S. M. Mohammad, X. Zhu, S. Kiritchenko and J. Martin, "Sentiment, emotion, purpose, and style in electoral tweets," *International Process Management*, vol. 51, no. 4, pp. 480–499, 2015.

[14] E. Cambria, "Affective computing and sentiment analysis," *IEEE Intelligence System*, vol. 31, no. 2, pp. 102–107, 2016.

[15] M. Jabreel, A. Moreno and A. Huertas, "Do local residents and visitors express the same sentiments on destinations through social media?" in *Information and Communication Technologies in Tourism*, vol. 1, New York: Springer, pp. 655–668, 2017.

[16] T. Rupa Rani, G. Lavanya, Dr. H. Shaheen, R. Anusha and K. Nartkannai, "Open IOT service platform technology with semantic web," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, pp. 17–21, 2020.

[17] C. Chang, M. Monselise and C. C. Yang, "What are people concerned about during the pandemic? Detecting evolving topics about COVID-19 from Twitter," *Journal Health Information Recourse*, vol. 1, pp. 17–28, 2021.

[18] M. Z. Asghar, A. Khan, S. Ahmad, M. Qasim and I. A. Khan, "Lexicon-enhanced sentiment analysis framework using rule-based classification scheme," *PLoS One*, vol. 12, no. 2, pp. 1–22, 2017.

[19] C. Dos Santos and M. Gatti, "Deep convolutional neural networks for sentiment analysis of short texts, cooling," in *25th Int. Conf. on Computational Linguistics*, Dublin, Ireland, vol. 1, pp. 125–139, 2014.

[20] X. Wang, W. Jiang and Z. Luo, "Combination of convolutional and recurrent neural network for sentiment analysis of short texts," in *COLING 2016, the 26th Int. Conf. on Computational Linguistics, December 20*, Osaka, vol. 1, pp. 895–910, 2016, December 20.

[21] X. Ouyang, P. Zhou, C. H. Li and L. Liu, "Sentiment analysis using convolutional neural network," in *IEEE Int. Conf. on Computer and Information Technology; Ubiquitous Computing and Communications Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing*, vol. 1, pp. 26–28, 2015.

[22] D. Zimbra, A. Abbasi, D. Zeng and H. Chen, "The state-of-the-art in Twitter sentiment analysis," *ACM Transaction Management Information System*, vol. 9, no. 2, pp. 1–2, 2018.

[23] S. He, Z. Li, Y. Tang, Z. Liao and F. Li, "Parameters compressing in deep learning," *Computers, Materials & Continua*, vol. 62, no. 1, pp. 321–336, 2020.

[24] S. R. Zhou and B. Tan, "Electrocardiogram soft computing using hybrid deep learning CNN-ELM," *Applied Soft Computing*, vol. 86, no. 4, pp. 105778, 2020.

[25] Y. Chen, H. Zhang, R. Liu, Z. Ye and J. Lin, "Experimental explorations on short text topic mining between LDA and NMF-based Schemes," *Knowledge-Based Systems*, vol. 1, no. 5, pp. 1–13, 2019.