



Adapting recurrent neural networks for classifying public discourse on COVID-19 symptoms in Twitter content

Samina Amin¹ · Abdullah Alharbi² · M. Irfan Uddin¹ · Hashem Alyami³

Accepted: 18 May 2022 / Published online: 10 August 2022

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

The COVID-19 infection, which began in December 2019, has claimed many lives and impacted all aspects of human life. With time, COVID-19 was identified as a pandemic outbreak by the World Health Organization (WHO), putting massive pressure on global health. During this ongoing pandemic, the exponential growth of social media platforms has provided valuable resources for distributing information, as well as a source for self-reported disease symptoms in public discourse. Therefore, there is an urgent need for effective approaches to detect self-reported symptoms or cases in social media content. In this study, we scrapped public discourse on COVID-19 symptoms in Twitter content. For this, we developed a huge dataset of COVID-19 self-reported symptoms and gold-annotated the tweets into four categories: confirmed, death, suspected, and recovered. Then, we use a machine and deep machine learning models, each with its own set of features, such as feature representation. Furthermore, the experimentations were achieved with recurrent neural networks (RNNs) variants and compared their performance with traditional machine learning algorithms. Experimental results report that optimizing the area under the curve (AUC) enhances model performance, and the long short-term memory (LSTM) has the highest accuracy in detecting COVID-19 symptoms in real-time public messaging. Thus, the LSTM classifier in the proposed pipeline achieves a classification accuracy of 90.7%, outperforming existing state-of-the-art algorithms for multi-class classification.

Keywords Deep learning · Coronavirus · Pandemic · COVID-19 · Classification · Recurrent neural networks · Twitter

Communicated by Shah Nazir.

✉ M. Irfan Uddin
irfanuddin@kust.edu.pk

Samina Amin
kustsameena@gmail.com

Abdullah Alharbi
amharbi@tu.edu.sa

Hashem Alyami
hyami@tu.edu.sa

¹ Institute of Computing, Kohat University of Science and Technology, Kohat 2600, Pakistan

² Department of Information Technology, College of Computers and Information Technology, Taif University, P. O. Box 11099, Taif 21944, Saudi Arabia

³ Department of Computer Science, College of Computers and Information Technology, Taif University, P. O. Box 11099, Taif 21944, Saudi Arabia

1 Introduction

SARS-CoV-2, a novel coronavirus, emerged in December 2019 to start a pandemic of respiratory illnesses identified as COVID-19, which has proven to be a difficult infection that can manifest in a variety of forms and severities ranging from minor to serious, with the risk of organ damage and death, from a moderate, self-limiting viral infection to serious, progressive pneumonia, multi-organ failure, and mortality (Chen et al. 2020; Wang et al. 2020; Huang et al. 2020). With the pandemic's emergence and the increasing number of reported cases and patients suffering from severe respiratory failure and cardiac disorders, there are valid reasons to be worried about the virus's implications (Guo et al. 2020). The importance of identifying effective solutions to solve COVID-19-related challenges has garnered considerable attention. However, another major issue that scientists and researchers face is the ever-increasing amount of information, also known as big data, which poses a challenge in the battle against novel coronavirus infection.

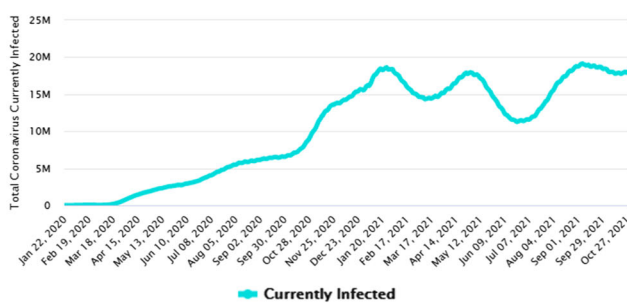
Social media (SM) has greatly grown into a vital communications platform for the collection, diffusion, and consumption of information since the outbreak of the COVID-19 pandemic. SM is becoming the go-to platform for public thoughts, attitudes, and impressions of COVID-19-related incidents and public health policy as a result of the epidemic outbreak (Pérez-Escoda et al. 2020). For government agencies, healthcare organizations, and institutions, SM has become a critical networking mechanism to disseminate critical awareness to the public.

SM can also be efficiently used to transfer health information to the general public during a pandemic outbreak. For example, several researchers have used SM data, particularly Twitter, to detect and identify the spread of the disease outbreaks as well as analyze public views, actions, and preconceptions (Amin et al. 2020). Other important studies employ Twitter data in combination with machine and deep learning techniques to detect and identify disease outbreaks during the Zika virus (Khatua et al. 2019; Ghenai and Mejova 2017), dengue (Amin et al. 2020), Ebola (Oyeyemi et al. 2014), and flu/influenza (Amin et al. 2020).

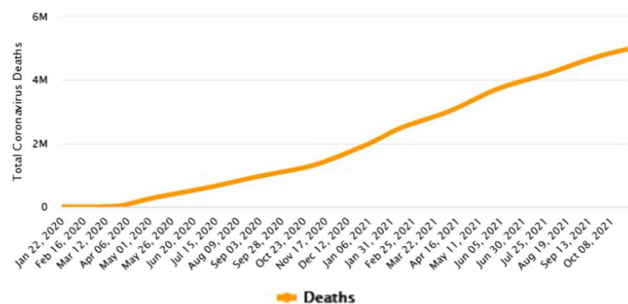
The coronavirus outbreak has now extended to almost the entire world after the COVID-19 appearance in Wuhan city in China, taking a massive hit on human lives (Chen et al.

2020). Although China's severity has declined substantially, many other countries, notably the USA and some in Europe, are still suffering from its influence. About 30.2 million people have been contaminated by COVID-19 till September 15, 2021, and the outbreak has taken 946 lacs of human lives worldwide, while 20.5 million have been recovered. Figure 1 (Worldometers 2004) shows the distribution of COVID-19 (a) "currently infected cases" or "active cases" worldwide from January 22, 2020, to September 29, 2021. The x-axis shows the flow of COVID-19 at a certain period, while the y-axis represents the number of COVID-19 confirmed cases in million, and Fig. 1b depicts the global distribution of COVID-19-related deaths from January 20, 2020, to April 15, 2021. Figure 1c shows the number of "recovered" or "cured" or "discharged" cases from COVID-9, while Fig. 1d shows the number of total "critical" or "serious" cases related to COVID-19.

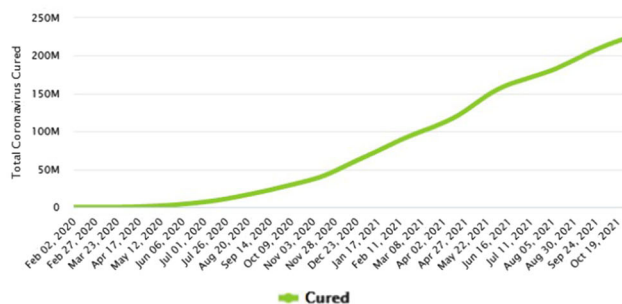
Twitter is among the most popular SM platforms around the world, and it has been used to spread information about the COVID-19 pandemic to provide early warnings. This demonstrates how and to what level artificial intelligence (AI) could be essential for the development and improvement of public healthcare systems (Hamet and Tremblay 2017). Due to this, there is a crucial need to build a model that can detect



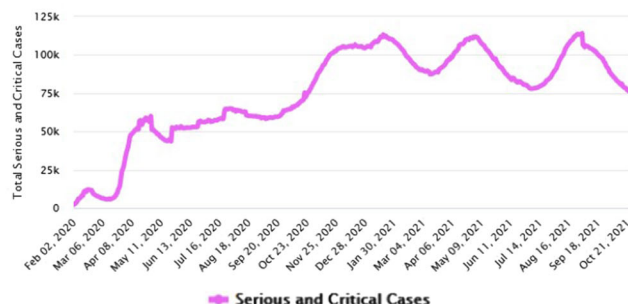
(a) Monthly wise distribution of total confirmed cases of COVID-19



(b) Monthly wise distribution of total deaths due to COVID-19



(c) Monthly wise distribution of total recovered or cured cases from COVID-19



(d) Monthly wise distribution of COVID-19 "critical" or "serious" cases

Fig. 1 Monthly wise distribution of total **a** confirmed cases, **b** deaths rates, **c** recovered cases, **d** "critical" or "serious" cases of COVID-19

the spread of the COVID-19 pandemic in real-time from Twitter messages automatically. In this work, we use natural language processing and deep learning models to detect the spread of the COVID-19 pandemic on Twitter. Specifically, LSTM (Sepp Hochreiter and Jurgen Schmidhuber 1997) with TF-IDF (Medina and Ramon 2003) and N-gram (Violos et al. 2018) is applied to record the long-term associations from the COVID-19 tweets; i.e., if a sequence is large enough, it will be difficult to convey information from earlier to later time steps. The main contributions of the proposed work are listed below.

- We derive a subset of tweets linked to the COVID-19 epidemic outbreak from a massive tweet corpus on COVID-19. Labeling the sample is done by human annotators. We can automatically evaluate the validity of the deemed tweets using high-quality, human-powered data labeling.
- For a comparison of the models, we develop the variants of RNNs using bidirectional recurrent neural networks (BiRNN), LSTM and gated recurrent units (GRUs) with TF-IDF and n-gram techniques based on more than four thousand tweets related to the COVID-19 pandemic.
- We evaluate the results of the proposed model with three conventional models with different features, including artificial neural network (ANN), logistic regression, naïve Bayes, support vector machine (SVM), and decision tree, as well as three deep learning classifiers such as BiRNN, LSTM, and GRU.
- We boost the performance of the selected methods by using accuracy, precision, recall, F-measure, confusion matrix, and the receiver operating characteristic (ROC) loss function to optimize the region under the curve by generating the true-positive rate (TPR) against the false-positive rate (FPR) at different parameter settings.

We organize our article into sections as follows: Section 2 deals with the literature review, and the proposed model is developed in Sect. 3. The achieved outcomes and their analysis are reported in Sect. 4, while Sect. 5 outlines the conclusion and future work.

2 Literature review

The affliction caused by COVID-19 is extensive (Holshue et al. 2020). People commonly utilize SM as a forum to post and share their ideas and sentiments. Scholars have been using SM data for public opinion polling and election prediction for years (Murthy et al. 2016), sentiment analysis (Mengistie and Kumar 2021), product promotion (Bollen et al. 2011), earthquake prediction (Earle et al. 2011; Hernandez-Suarez et al. 2019), disruptive event prediction

(Alsaedi et al. 2017), the spread of misinformation (Ferrara et al. 2020), public health monitoring (Amin et al. 2020, 2021), and so forth. Twitter is an SM application that allows users to post text, images, and short videos (Gleason 2018). To collect data for this study, we used Twitter as an SM platform.

In the field of a pandemic outbreak, one of the main areas of concentration is emergency communication and collaboration. Possibly, this is because SM provides timely outlets for government institutions and healthcare responders to engage with citizens and for people to disseminate such information with their relatives and friends. This is significant when it comes to pandemic outbreaks. Yang and Sun (Yang and Su 2020) looked into COVID-19's health policy and examined how the function of SM in promoting policy evolution can aid healthcare responders and government. Abrams and Greenhawt (2020) manifested that using SM platforms to offer a continuing and regular communication way is one strategy to maintain proper risk communication throughout the COVID-19 pandemic. During the 2019 measles outbreak in the USA, (Kim and Hawkins 2020) stated that SM could help raise awareness and encourage positive health prevention behaviors. In particular, SM communication and transmission can aid in the promotion of preventive personal grooming intentions. Cotfas et al. (2021) examined the dynamics of public opinion on COVID-19 vaccination over one month beginning with the first vaccine declaration and ending with the first vaccination in the UK, during which civil society expressed a keen interest in the vaccination procedure. The best performing classifier was chosen by comparing traditional machine learning and deep learning techniques. Ferrara et al. (2020) analyzed the spread of misinformation and abuse on the COVID-19 pandemic on social media. Iqbal et al. (2021) used LSTM to forecast the number of COVID-19 patients in Pakistan. In their work, they used COVID-19 data from Pakistan (March 2020 to May 2020) to train the RNN model and forecast the COVID-19 percentage of positive patients for June 2020.

Twitter data have been retrieved in a number of studies since the COVID-19 pandemic to clarify public responses and dialogues about COVID-19 (Chun et al. 2020; Ahmed et al. 2020; Sahni and Sharma 2020; Hung et al. 2020; Kruspe et al. 2020). Chun et al. (2020) investigated how citizens' attitudes changed throughout the COVID-19 outbreak. The authors provided a method for measuring and tracking citizens' levels of concern by assessing sentiment in Twitter data and dividing the total number of tweets by the ratio of negative and highly negative tweet counts. They looked at 30,000 tweets from March 14, 2020, to see the level of concern across the US state. Mukhtar (2020) emphasized the necessity of addressing the mental health and behavioral issues that led to the COVID-19 pandemic. They said that mental health and well-being are critical components of

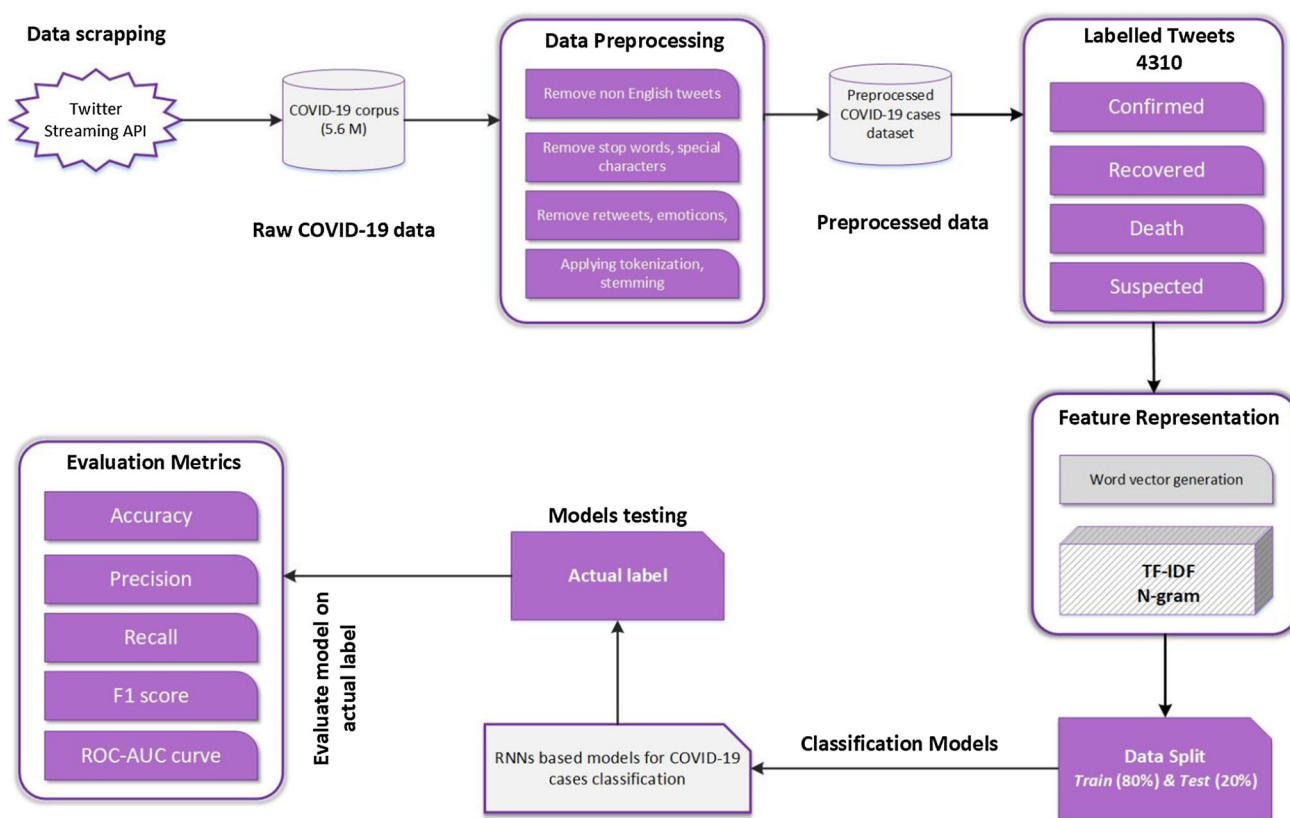


Fig. 2 Proposed system design

healthcare and that understanding and addressing these concerns are critical for a stable and healthy community. They also looked into some of the factors that could lead to mental health problems, such as the unknown nature of this novel disease; the unpredictability of emerging threats such as self-isolation, social distancing, and quarantine; impairments in social functioning; interpersonal issues; the persistence of emotional and behavioral disorders and psychological problems; preconditioned mental health issues; and the tendency to be easily influenced by traumatic incidents. Amin et al. (2021) introduced a machine learning-based model to detect seasonal outbreaks using Twitter data.

Moreover, (Zhang et al. 2020) proposed tracking depression trends via Twitter during the COVID-19 outbreak. The authors compiled a large English Twitter depression dataset, which included 2575 users and their tweets from history. They used three categorization models to look at how depression affected people's Twitter language. To analyze these features with depression signals, deep-learning model scores were merged with psychological text features and user demographic information. Similarly, (Galbraith et al. 2021) developed a model based on Twitter data to better understand the thinking of Indians during the lockdown. By performing

analysis, they employed Python and R statistical software and aimed to generate participation along two interconnected lines, in particular, to maintain the integrity of online discussions on social media. Another motivated work was carried out in Li et al. (2021), Amin et al. (2021) by exploring the potential of SM data to monitor early warning of the COVID-19 outbreak.

To conclude, decision-makers and healthcare organizations will benefit greatly from establishing a detection technique based on SM content. With fewer contributions to COVID-19 infection detection, the majority of the existing work has been done for fake news detection or sentiment analysis. The detection and monitoring of epidemic outbreaks over time were not a priority of the research studies on self-reported cases or symptoms in public discourse. That inspired us to conduct our study on Twitter data in order to better understand COVID-19 symptoms in public discourse during the pandemic to help healthcare responders, governments, and decision-makers. To the best of our knowledge, we are the first to explore the potential of an automated model based on RNNs with TF-IDF and N-grams for detecting and classifying public discourse on COVID-19 symptoms cases in tweets.

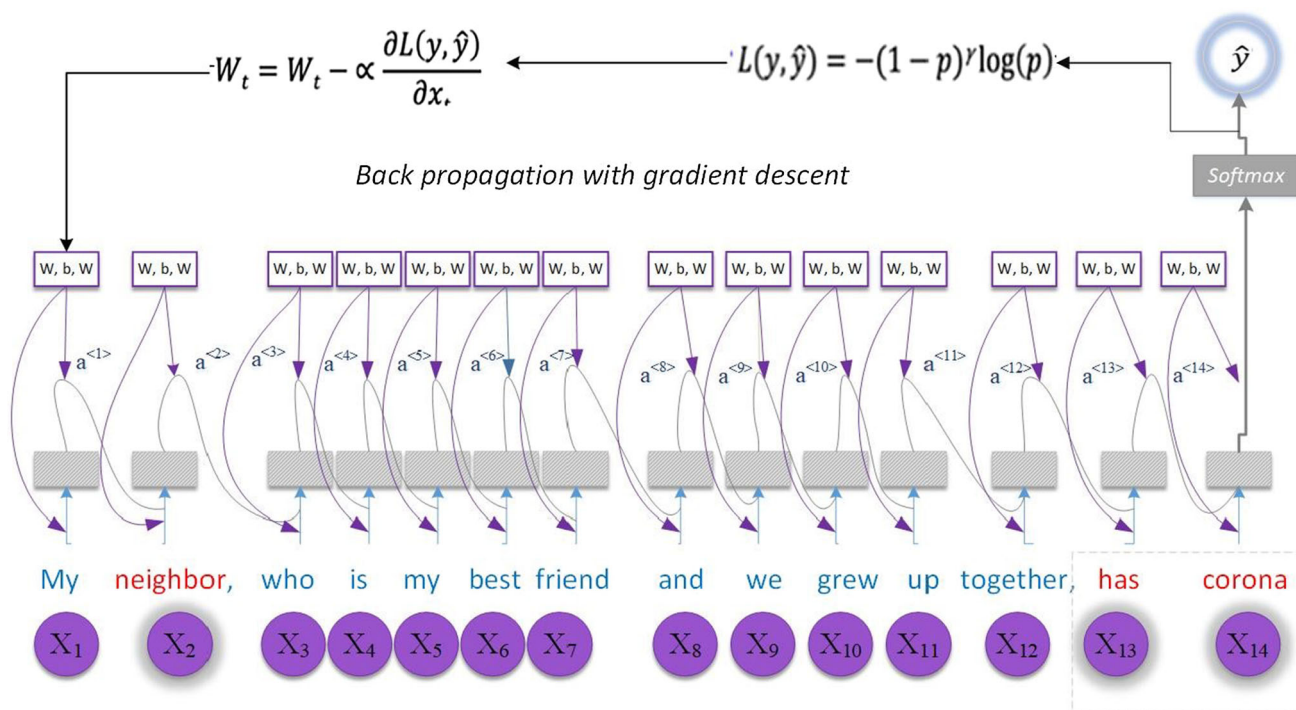


Fig. 3 RNN to process long sentence

3 Materials and methods

In this section, the working pipeline of the proposed models is demonstrated. Figure 3 shows the overall methodology. It starts by gathering data with the help of the Twitter streaming API and finishes by evaluating and comparing the performance of the proposed models. These steps are described in detail as follows.

3.1 Data gathering

The information that people choose to disclose in public is reflected in Twitter messages. In this work, we used the Twitter Standard Search API (Twitter scraper 2018) to download linked tweets from January 20, 2020, to April 15, 2021, using the search phrase “Coronavirus,” “COVID-19” Additionally, we chose the keywords including “#COVID-19,” “#corona,” “#covid19,” “#coronavirus,” and “#covid” to gather tweets because it may create enough data reflective of popular opinion. We used the API to gather related tweets daily from the beginning of the study period to confirm the reliability of the data linked with the target topic. The data were scraped on a daily basis and recorded in simple JavaScript Object Notation (.JSON) files and converted to comma-separated value (.CSV) files for further processing.

3.2 Data wrangling

The dataset is then normalized and cleaned by applying the natural language toolkit (NLTK)¹ python libraries in NLP (Bird et al. 2009). We used multiple techniques to clean the tweets, as shown in “Data Preprocessing” in Fig. 1. In a tweet, short URLs, @username, RT @username, punctuations, numerals, emojis, and stop words were eliminated. Each tweet was then tokenized or broken down into a list of distinct words. The tokenized words were transformed to their stemmed forms because the words in a tweet were written in different ways. This process is termed as lemmatization.

3.3 Data labeling

To develop the training dataset, we took the top 4,010 most popular tweets and manually categorized each one as “confirmed,” “death,” “recovered,” and “suspected.” Three annotators label the sample tweets in order to eliminate gaps or preconceptions in the annotation. In this work, a “confirmed” tweet denotes that when someone is infected with severe COVID-19 symptoms, a class confirmed is given. For example, all tweets that show that someone is infected with COVID-19 (such as a person who has corona) are consid-

¹ https://www.nltk.org/_modules/nltk/stem/porter.

Table 1 Examples of COVID-19 confirmed, death, suspected, and recovered cases in tweets

Tag	Tweet
<i>confirmed</i>	This is sheer evil Remember, that till 14th Sept, 18102 kids were infected with #COVID19, and ZERO died. #covidinfected I am currently infected with covid 19 right now my body feels terrible. But I'm trying my best to recover soon #COVID19 I am again seeing whole families infected with #COVID19. I don't want to do this again. #GetVaccinated #WearAMask
<i>death</i>	TRAGIC: 41-yo elementary schoolteacher Kelly Peterson died of #COVID-19 It is informed with grief that Brig Nadeem, 95 PMA (ex DA UK) died due to Covid19/Lungs failure. This resulted in Cardiac arrest. May Allah bless him in the highest place in Jannah RIP - #COVID19 has claimed the life of another South Florida police officer. Sergeant Patrick "Pat" Madison died on Friday due to complications of COVID-19, according to the police department
<i>recovered</i>	Very happy to share that both of my parents and my brother recovered from rather severe cases of #COVID19 and doing great with all your prayers and love I along with Dad and Mom have won the battle against #COVID19. We are now fully recovered from #COVID19. The last 17 days have been life-changing experiences for me. I am a different person all together learned many thing's. Keep Faith in God
<i>suspected</i>	I got infected with Covid after 1st shot of the vaccine. But due to 1st shot of the vaccine, the covid load was mild I am at risk of being infected with Covid19...I knew this like a few days ago Well, I have had all the symptoms of COVID-19 but took 3 tests and they were all negative moral of the story: go get vaccinated It is frustrating to miss work when you have a mild cold that typically you would tough out and go to work for. I got my negative Covid results 3 days ago but need to be symptom-free for 24 hours and this congestion and slight cough are lingering. had to leave work early today

ered confirmed, while the suspected tweet implies that the people thought the corona infection was likely to happen in them, possibly due to the symptoms of COVID-19 (e.g., a tweet depicting mild symptoms of COVID-19 with the term "suspected" attached). A label death is assigned to a tweet that shows the death caused by COVID-19 (e.g., a tweet depicting a death caused by COVID-19 with the label "death" attached). More information about manually labeled tweets belonging to the four classes is given in Table 1.

To ensure the quality of the annotated COVID-19 dataset that would be employed for the proposed models, retweets and duplicates were removed. Three independent raters classified the tag of the tweets on COVID-19 symptoms into four classes, namely: "confirmed," "death," "recovered," and "suspected." Disagreement has only been captured between confirmed and suspected or between the suspected and recovered tags. There has been no disagreement between the annotations "confirmed and death," or "recovered and death," and "suspected and death." In the case of a disagreement, the tweet has been assigned to the class approved by the majority of human raters. The statistical distribution of tweets in the tagged dataset in the four considered classes is given in Table 2.

Table 2 Statistics for the manually tagged COVID-19 dataset

Class	Tweet#	Tweet%
<i>confirmed</i>	1662	38.56%
<i>death</i>	909	21.09%
<i>recovered</i>	1053	24.43%
<i>suspected</i>	686	15.92%
Total	4,310	100.00%

3.4 Feature representation using TF-IDF

Text content must first be transformed into numerical feature vectors before machine learning methods can be used to classify it. The bag-of-words (BoW) method transforms text to numeric values using the frequency of the words as a preliminary step. Assume a vocabulary $V=w_1, \dots, w_N$, comprising N tokens, represented using w_i , a tweet, a document d , belonging to a COVID-19 corpus $C_{covid} - 19$, can be depicted using a feature vector $X=x_1, \dots, x_N$, in which x_i can be a binary variable showing whether or not the word w_i occurs in the $C_{covid} - 19$ corpus, or a numeric variable showing the number of times the term w_i comes in the text.

Since frequently occurring words can sometimes have little “relevant information,” classification approaches that depend on word frequencies can benefit from a more complex feature representation called term frequency-inverse document frequency (TF-IDF), which minimizes the weight attached to words that appear often throughout the corpus. The TF-IDF in (1) and (2) is calculated as follows:

$$TF(w, d) = \log(1 + f(w, d)) \quad (1)$$

$$IDF(w) = \log\left(\frac{1 + N}{1 + n_w}\right) \quad (2)$$

where TF (t,d) shows the count of word w in document d. n shows the number of documents that consist only of the word w, and N is the total number of documents. For feature representation, the TF-IDF statistical method is applied through the present work.

The BoW technique does not give any information about the order of the words because it just looks at how many times they appear in a text. This problem can be solved by employing the N-gram language model. N-gram has been proven to outperform other commonly used models. In N-grams, the text is depicted by a series of n-sequential words. Common variants of N-grams are unigrams (1-gram), bigrams (2-gram), and trigrams (3-gram). In the current study, three variants of N-grams are applied and promising results are obtained with the bigram model. When utilizing the TF-IDF bigram, all of the classifiers achieved the best results.

3.5 Deep learning algorithms

After vector generation, the next step is to develop machine and deep learning algorithms. The proposed solution is carried out on real-time COVID-19 data from publicly available users’ conversations on Twitter, which is made available by an open-source Twitter scraping API (see Sect. 3.1).

3.5.1 RNN, LSTM, and GRU

RNN is a neural network that deals with data in sequential order (Lipton et al. 2015). Sequential data (also known as time series) can take the shape of text, video, speech recognition, voice recognition, time series prediction, NLP tasks, and other media. RNN generates the current output by using the prior information in the sequence. If a sequence is large enough, it will be difficult to convey information from earlier to later time steps. For example, RNN may omit significant information from the beginning while processing a paragraph of text to make a prediction as shown in Fig. 4. Internally, it has very few operations, but when the conditions are correct, it performs quite well (like short sequences). However, LSTM and GRU were developed as the alternative to short-term memory or vanishing gradient problems in RNN.

The GRU (Chung et al. 2014) is an advanced variant of RNN that looks a lot like an LSTM. The GRU did away with the cell state and is now using the invisible state to move data. It has two gates: a reset gate and an update gate. The update gate functions similarly to an LSTM’s forget and input gates. It determines what information should be discarded and what should be added. However, the reset gate was used to determine how much previous information to overlook. Since GRU has fewer tensor operations, they are somewhat faster to train than LSTM.

The control flow of an LSTM (Sepp Hochreiter and Jurgen Schmidhuber 1997) is similar to that of an RNN. It analyzes the data and passes the information on as it moves ahead. The operations within the cells of the LSTM vary. The LSTM uses these operations to remember or forget information. There are two additional gates here in addition to GRU. 1) output gate and 2) forget gate. Here, the forget gate regulates what is remembered and what is forgotten from a previous cell state. The input gate determines which elements of the cell are output to the hidden state, while the output gate regulates how much information from the previous state should be reserved or forgotten. It will determine what the next hidden state will be. Thus, the LSTM cell uses a forget gate to perform element-wise multiplication on the prior memory cell state.

3.5.2 Parameters settings for simulations

The proposed model is developed in Python (Anaconda 2020) and is based on the machine and deep learning open source libraries; (Pedregosa et al. 2011; Chollet 2015; Team 2015) and (Oliphant 2006), and it runs on a 64-bit OS with 8GB RAM and an Intel Core i5 CPU. The COVID-19 dataset (Table 2) is used to design and simulate the proposed model. The dataset is first preprocessed for simulations using the NLTK libraries (see Sect. 3.2). To train the model, the pre-processed dataset is split into training, and testing sets during the training process, with a training ratio of 80% and testing proportions of 20%. However, the LSTM model is made up of three layers, each of which comprises 256, 128, 64, and 4 neurons and includes batch size, learning rate (0.01), and dropout layers. Furthermore, the BiRNN and GUR models’ configurations are identical to those mentioned above. The training of LSTM is first performed by running 10, 15, and 30 iterations with a batch size of 32 using the Adam optimizer, whereas the training of BiRNN and GRU is primarily conducted by performing 20 training iterations with a batch size of 32 using the Adam optimizer. Besides, an optimization technique is used to select the optimal hyperparameters for the proposed and other traditional models. To find an optimal method, Table 3 lists the selection of the optimal hyperparameters for the proposed model, whereas the pseudocode used for the proposed model is provided in Algorithm 1.

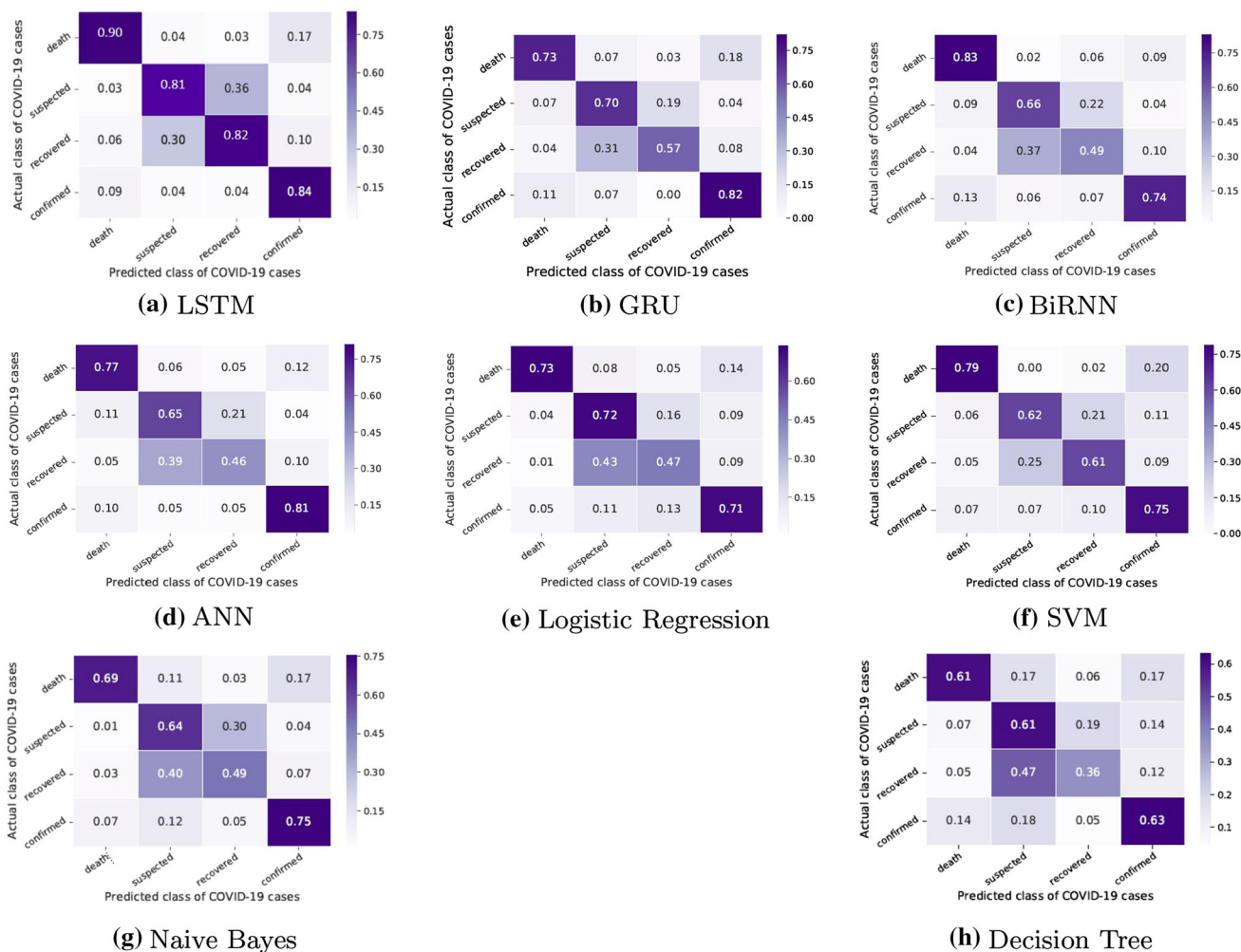


Fig. 4 The confusion matrix of the models classifying confirmed, death, recovered, and suspected cases in tweets

Table 3 Optimal hyperparameters settings for f the proposed RNNs model

Parameters	BiRNN	LSTM	GRU
Pre-trained vocabulary size	23000	23000	23000
ngram_range	(1, 2) (1, 3)	(1, 2) (1, 3)	(1, 2) (1, 3)
Number of hidden layers	3	3	3
Number of neurons in hidden layers	256, 128, 64	256, 128, 64	256, 128, 64
Output layer	1	1	1
Number of neurons in the output layer	4	4	4
Learning rate	0.001	0.001	0.001
Optimizer	adam	adam	adam
Epoch no#	10, 15, 30	10, 20, 40, 50	10, 20, 40, 50
Batch size	32	32	32
Activation	softmax	softmax	softmax
Loss function	binary_crossentropy	binary_crossentropy	binary_crossentropy
Dropout	0.2, 0.5	0.2, 0.3, 0.5	0.2, 0.5
Training time (sec)	64	68	62
Test loss	0.88	0.23	0.57

Algorithm 1 Classifying COVID-19 symptoms into cases: confirmed, death, recovered, and suspected

```

Begin
Input: COVID-19 tweet data
Output: Classifying COVID-19 symptoms into confirmed, death,
recovered, and suspected
Parameters: COVID-19 Tweet  $\leftarrow T_{covid-19}$ , COVID-19 Corpus
 $\leftarrow C_{covid-19}$ , Re-Tweets  $\leftarrow R_{Tw}$ , Remove Re-Tweets:  $R_{RTw}$ ,
Parse Tweet File to CSV  $\leftarrow PTF\_to\_CSV$ , Remove Username
 $\leftarrow R_{UserNm}$ , Remove URL  $\leftarrow R_{URL}$ , Remove Tag#  $\leftarrow R_{\#}$ ,
Word  $\leftarrow w$ , Regular Expression  $\leftarrow R_{Exp}$ ,  $RemoveStopWords$ 
( $a, is, be, the, etc.$ )  $\leftarrow R_{SpW}$ 
Set:  $PTF\_to\_CSV \leftarrow [null]$ 

Begin
Initialization set of keywords used to fetch  $KW_{covid-19}$ 
Build a JSON file to save  $T_{covid-19}$ 
Access credentials (IDs) of Twitter scraping API
begin
for each  $T_{covid-19}$  do
  search ( $T_{covid-19}$ ,  $KW_{covid-19}$ )
  if ( $T_{covid-19}$  relevant to  $KW_{covid-19}$ ) then
    save_to_  $C_{covid-19}$  ( $C_{covid-19}$ )
  else
    discard ( $T_{covid-19}$ )
  end if
end for
end
for each  $T_{covid-19}$  in  $C_{covid-19}$  do
  covid19_dataset  $\leftarrow$  json.load ( $T_{covid-19}$ )
  PTF_to_CSV  $\leftarrow$  covid19_dataset ['Text']
  save_to_csv_file ( $PTF\_to\_CSV$ )
end for
for each  $T_{covid-19}$  in  $PTF$  do
  language  $\leftarrow$  check_language ( $T_w$ )
  if (language  $\leftarrow$  "english") then
    save_to_  $C_{covid-19}$  ( $T_{covid-19}$ )
  else
    discard_from_  $C_{covid-19}$  ( $T_{covid-19}$ )
  end if
end for
repeat
  preprocess
   $R_{RTw} \leftarrow R_{Exp}$ .findall ( $T_{covid-19}$ , "RT W@w+")
   $R_{UserNm} \leftarrow R_{Exp}$ .findall ( $T_{covid-19}$ , "@/w+")
   $R_{URL} \leftarrow R_{Exp}$ .findall ( $T_{covid-19}$ , "http(S)?://(. \| [0-9] \| [a-z]
\| [A-Z] +)")
   $R_{\#} \leftarrow R_{Exp}$ .findall ( $T_{covid-19}$ , "#/w+")
   $R_{SpW} \leftarrow R_{Exp}$ .findall ( $T_{covid-19}$ , "[a, be, the, is, if]/w+")
   $R_{RTW} \leftarrow R_{UserNm} = R_{URL} = R_{\#} = R_{SpW} = [NULL]$ 
until  $T_{covid-19} \leq$  empty  $C_{covid-19}$ 
do perform tokenization and stemming
do generate w vector using  $TF - IDF$  and ngram_range
do split  $C_{covid-19}$  into a train(80%) and test (20%) sets
do find optimal model perform classification
max feature(TF-IDF)  $\leftarrow$  7500, ngram_range  $\leftarrow$  (1, 2) (1, 3), number
of hidden node  $\leftarrow$  3, dropout  $\leftarrow$  0.5, epochs  $\leftarrow$  30, batch size: 32
for test set do
  compute performance metrics
  accuracy, precision, recall, F1-score, ROC-AUC
  evaluate a model on actual  $C_{covid-19}$ 
end for
End

```

3.5.3 Performance metrics for accuracy evaluation

The proposed RNNs-based models' accuracy is evaluated through well-known metrics (Powers 2020): precision, recall, F1-score, and ROC curve. These performance metrics are calculated from the confusion matrix (such as a matrix that produces different outcomes in classification tasks). The confusion matrix generates four rows and four columns for the proposed classification task, i.e., eight alternative outcomes. The performance measures given in (3 - 7), as defined in Powers (2020), are as follows.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (3)$$

$$Precision = \frac{(TP)}{(TP + FN)} \quad (4)$$

$$Recall = \frac{(TP)}{(TP + FP)} \quad (5)$$

$$F1 - score = \frac{(Precision \times Recall)}{(Precision + Recall)} \quad (6)$$

$$AUC = \sum_{i \in (TP + FP + TN + FN)} \frac{(TPR_i + TFR_{(i-1)}) \times (FPR_i + FPR_{(i-1)})}{2} \quad (7)$$

A classifier's accuracy is a measure that shows the percentage of correct predictions. A recall is also referred to as sensitivity, and true-positive rate (TPR) is another class imbalance metric. It demonstrates how a tweet can detect COVID-19 symptoms. Similarly, precision refers to a classifier's ability to correctly classify COVID-19 symptoms. However, using the imbalanced dataset, accuracy, precision, and recall metrics cannot generate an accurate picture of the model's efficiency (Amin et al. 2021; Powers 2020). In comparison with the metrics mentioned above for performance evaluation, the F1-score is another helpful class imbalance measure. The F1-score provides a good combo of precision and recall, demonstrating its worth in comparison with other metrics.

Moreover, the AUC (Hajian-Tilaki 2013) is a performance metric that is more dependable and accurate for an unbalanced dataset. Over the unbalanced dataset, it provides a more accurate evaluation of the model's performance in terms of COVID-19 cases detection. It is the probability that the model will rank positive data better than negative data. AUC is also known as the area under the receiver operating characteristic curve (ROC-AUC), which is used to evaluate the model's ability to distinguish between classes. The ROC-AUC is a graphical depiction that plots the TPR against the FPR to evaluate the proposed model performance. Furthermore, the

Table 4 Comparison of the proposed model performance and all the baselines on the TF-IDF feature representation

Model	Accuracy (%)	Precision	Recall	F-score
LSTM	90.7	0.89	0.90	0.89
GRU	86.6	0.86	0.85	0.87
BiRNN	86.1	0.85	0.86	0.86
ANN	83.2	0.82	0.83	0.82
SVM	76.7	0.77	0.76	0.76
Logistic Regression	76.6	0.77	0.78	0.77
Naïve Bayes	75.4	0.75	0.76	0.74
Decision Tree	73.2	0.72	0.71	0.72

area under the ROC curve is likely to be between 0 and 1. If a classifier's ROC-AUC score is more than 0.5, it gives a better detection rate when compared to random predictions, whereas if the classifier's ROC-AUC is less than 0.5, it means the classifier has poor classification performance.

4 Results and discussion

In this section, we present the experimental results of the proposed model as well as a comparison of its performance with state-of-the-art algorithms. The proposed solution is carried out on real-time COVID-19 data from publicly available user conversations on Twitter. Furthermore, to validate the performance of the proposed RNNs based models, five performance evaluation metrics are employed to demonstrate the efficiency of the LSTM model over baseline techniques for COVID-19 symptoms classification in tweet messages.

Table 3 lists a comparison of the proposed model performance and all the baselines with TF-IDF embedding. We compare our proposed model to state-of-the-art algorithms as well as conventional classification approaches. For a fair comparison, we applied an optimization technique to determine the best relevant hyperparameters for the benchmark algorithms. It has been observed that the LSTM performed best when using the TF-IDF technique. The model was able to classify COVID-19 symptoms in cases with 90.7% accuracy, 0.89 precision, 0.90 recall, and an F1-score of 0.89. As shown in Table 4, the proposed model has the greatest accuracy score when compared to other benchmark models.

The direct comparison between the models and LSTM revealed that the performance gap mostly hinges on suspected, which is more heavily misclassified by the other. To address this gap, we examined at the confusion matrix,

which revealed that suspected is more frequently confused with recovered and confirmed. Figure 5 depicts the confusion matrix of the proposed models.

We used bigram TF-IDF feature representations to train the proposed model. The LSTM classifier with bigram feature representations obtained the best accuracy of 90.7%, AUC score of 81%, and F-measure of 0.89 with bigrams of all the classifiers, while the GRU classifier with bigram feature representations obtained the accuracy score of 86.6% and the best AUC score of 80%. The findings show that the size of the N-gram has an impact on the accuracy rate of different classifiers. When utilizing the TF-IDF bigram, all of the classifiers achieved the best results. The proposed model outperformed other models in terms of ROC-AUC. Figure 6 shows the ROC-AUC curve generated by the proposed classifiers using both TF-IDF and bigram approaches.

To sum up, the proposed model outperformed other models in terms of ROC-AUC, as depicted in Fig. 6. As shown in Table 4, our proposed model has the greatest accuracy score when compared to other benchmark models. The LSTM performs the best among traditional classifiers, with a micro-average ROC curve score of 81% and macro-average ROC curve score of 81%, a precision of 0.89, recall of 0.90, and an F1-score of 0.89, indicating that the LSTM classifier's accuracy is substantially better than all other classifiers. The GRU classifier came in second with an accuracy of 86.1%, and a 0.87 F1-score. Furthermore, in terms of ROC-AUC and accuracy, the proposed technique has covered more areas in comparison with the training and validation sets, demonstrating its superiority. As previously said, the fundamental goal of this study is to increase the accuracy and ROC-AUC area while minimizing the loss value, which our model has accomplished when compared to other models.

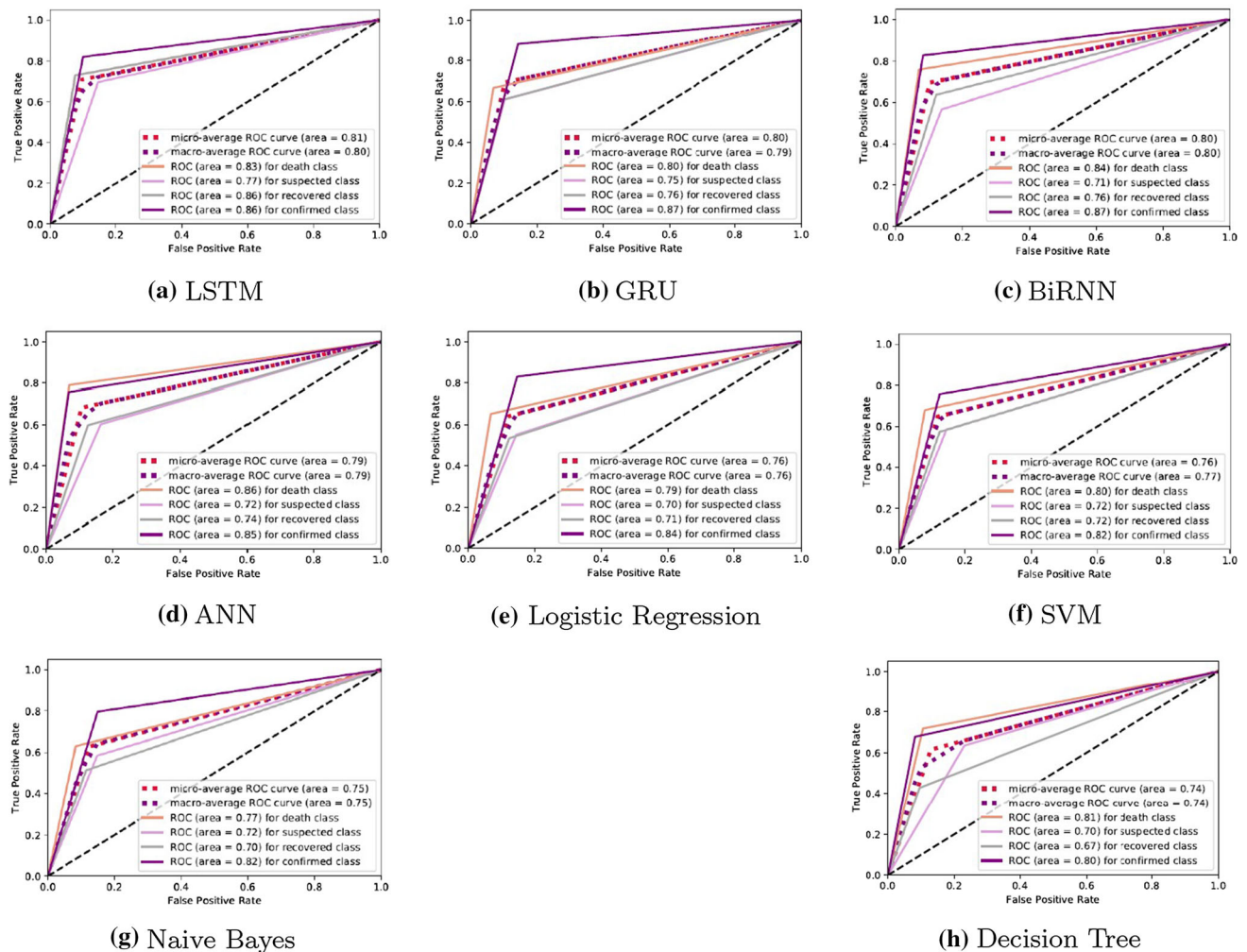


Fig. 5 ROC-AUC to evaluate the proposed models' performance by plotting TPR against the FPR and classifying confirmed, death, recovered, and suspected symptoms cases

5 Conclusion and future work

During the COVID-19 ongoing pandemic, the exponential growth of SM platforms has provided valuable resources for distributing information, as well as a source for self-reported disease symptoms in public discourse. Therefore, there is an urgent need for effective approaches to detect self-reported symptoms or cases in SM content. In this study, we proposed a detection and classification approach for the self-reported symptoms posted during the COVID-19 pandemic. For this, Twitter is used as a medium to collect more than five million tweets spanning from January 20, 2020, to April 15, 2021. Each tweet in the dataset was tagged with the case categories namely: confirmed, death, recovered, and suspected. We used this tagged dataset to develop machine and deep learning models and used an LSTM model to validate the dataset's quality. After annotation, we developed a detection and classification approach to track the self-reported

symptoms and cases of people to monitor and analyze the COVID-19 pandemic and make a classification report. For a fair comparison, we applied an optimization technique to find the optimal hyperparameters of the benchmark algorithms. It has been observed that the LSTM performed best when using the TF-IDF technique. The model was able to classify COVID-19 symptoms in cases with 90.7% accuracy, 0.89 precision, 0.90 recall, and an F1-score of 0.89. Our results assist the COVID-19 cases mentioned by the public and can be used to track early warning and help healthcare responders and government authorities. We compared our proposed approach to state-of-the-art algorithms as well as conventional classification approaches.

However, the proposed model has not been applied yet, to evaluate its efficiency on the advanced word embedding methods. The most important finding here is to explore such strong techniques based on features from the COVID-19 dataset. In the future, as symptoms categories are limited to

four classes, we are planning to expand the symptoms classes by including more cases categories. We also plan to evaluate the performance of advanced word embedding methods such as Word2Vec, Fasttext, Doc2Vec, and Glove to learn the semantic association among different words in SM text for better exploration. In addition, we aim to develop a web-based application that will crawl and label tweets in real time. The application will be useful in tracking user self-reported conversations about diseases in the course of any future pandemic outbreak.

Author contributions S.A. planned and conducted the review, designed the study, analyzed the results, and drafted the manuscript. A.A. provided the funding for the work and reviewing/editing the manuscript. M.I.U. contributed to planning, supervision of the work, and reviewing/editing of the manuscript. H.A. participated in reviewing data collection and the design of the study. All authors read and approved the final manuscript.

Funding This research was supported by Taif University Researchers Supporting Project number (TURSP-2020/231), Taif University, Taif, Saudi Arabia.

Data availability Enquiries about data availability should be directed to the authors.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

- Abrams EM, Greenhawt M (2020) Risk communication during covid-19, *The Journal of Allergy and Clinical Immunology. In Practice* 8(6):1791
- Ahmed W, Vidal-Alaball J, Downing J, Seguí FL et al (2020) Covid-19 and the 5g conspiracy theory: social network analysis of twitter data. *Journal of medical internet research* 22(5):e19458. <https://doi.org/10.2196/19458>
- Alsaedi N, Burnap P, Rana O (2017) Can we predict a riot? disruptive event detection using twitter. *ACM Transactions on Internet Technology (TOIT)* 17(2):1. <https://doi.org/10.1145/2996183>
- Amin S, Irfan Uddin M, Ali Zeb M, Alarood AA, Mahmoud M, Alkinani MH (2020) Detecting Dengue/Flu Infections based on tweets using LSTM and word embedding. *IEEE Access* 8:189054. <https://doi.org/10.1109/access.2020.3031174>
- Amin S, Uddin MI, Hassan S, Khan A, Nasser N, Alharbi A, Alyami H (2020) Recurrent neural networks with tf-idf embedding technique for detection and classification in tweets of dengue disease. *IEEE Access* 8:131522. <https://doi.org/10.1109/access.2020.3009058>
- Amin S, Alouffi B, Uddin M.I, Alosaimi W (2022) Optimizing convolutional neural networks with transfer learning for making classification report in covid-19 chest x-rays scans, *Scientific Programming* 2022
- Amin S, Uddin MI, Al-Baity HH, Zeb MA, Khan MA (2021) Machine learning approach for covid-19 detection on twitter, *Computers, Materials and Continua* pp. 2231–2247. <https://doi.org/10.32604/cmc.2021.016896>.
- Amin S, Uddin MI, AlSaeed DH, Khan A, Adnan A (2021) Early detection of seasonal outbreaks from twitter data using machine learning approaches, *Complexity* 2021. <https://doi.org/10.1155/2021/5520366>.
- Amin S, Uddin MI, Zeb MA, Alarood AA, Mahmoud M, Alkinani MH (2021) Detecting information on the spread of dengue on twitter using artificial neural networks, *CMC-COMPUTERS MATERIALS & CONTINUA* 67(1), 1317. <https://doi.org/10.32604/cmc.2021.014733>
- Anaconda software distribution (2020). <https://docs.anaconda.com/>
- Bird S, Klein E, Loper E (2009) *Natural language processing with Python: analyzing text with the natural language toolkit*. (O'Reilly Media,, 2009)
- Bollen J, Mao H, Zeng X (2011) Twitter mood predicts the stock market. *Journal of computational science* 2(1):1. <https://doi.org/10.1016/j.jocs.2010.12.007>
- Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, Qiu Y, Wang J, Liu Y, Wei Y et al (2020) Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *The Lancet* 395(10223):507. [https://doi.org/10.1016/S0140-6736\(20\)30211-7](https://doi.org/10.1016/S0140-6736(20)30211-7)
- Chollet F. Keras. <https://keras.io/api/layers/initializers/> (2015). [Online; accessed 19-July-2020]
- Chun SA, Li ACY, Toliyat A, Geller J (2020) in *The 21st Annual International Conference on Digital Government Research*, pp. 322–323
- Chung J, Gulcehre C, Cho K, Bengio K (2014) Empirical evaluation of gated recurrent neural networks on sequence modeling, arXiv preprint [arXiv:1412.3555](https://arxiv.org/abs/1412.3555)
- Cotfas LA, Delcea C, Roxin I, Ioanăș C, Gherai DS, Tajariol F (2021) The longest month: Analyzing covid-19 vaccination opinions dynamics from tweets in the month following the first vaccine announcement. *IEEE Access* 9:33203. <https://doi.org/10.1109/ACCESS.2021.3059821>
- Earle PS, Bowden DC, Guy M (2011) Twitter earthquake detection: earthquake monitoring in a social world, *Annals of geophysics* 54(6). <https://doi.org/10.4401/ag-5364>.
- Ferrara E, Cresci S, Luceri L (2020) Misinformation, manipulation, and abuse on social media in the era of covid-19. *Journal of Computational Social Science* 3(2):271. <https://doi.org/10.1007/s42001-020-00094-5>
- Galbraith N, Boyda D, McFeeters D, Hassan T (2021) The mental health of doctors during the covid-19 pandemic. *BJPsych bulletin* 45(2):93
- Ghenai A, Mejova Y (2017) Catching zika fever: Application of crowdsourcing and machine learning for tracking health misinformation on twitter, arXiv preprint [arXiv:1707.03778](https://arxiv.org/abs/1707.03778)
- Gleason B (2018) Thinking in hashtags: exploring teenagers' new literacies practices on twitter. *Learning, Media and Technology* 43(2):165. <https://doi.org/10.1080/17439884.2018.1462207>
- Guo T, Fan Y, Chen M, Wu X, Zhang L, He T, Wang H, Wan J, Wang X, Lu Z (2020) Cardiovascular implications of fatal outcomes of patients with coronavirus disease 2019 (Covid-19). *JAMA Cardiol* 5(7):811. <https://doi.org/10.1001/jamacardio.2020.1017>
- Hajian-Tilaki K (2013) Receiver operating characteristic (roc) curve analysis for medical diagnostic test evaluation. *Caspian J Int Med* 4(2):627
- Hamet P, Tremblay J (2017) Artificial intelligence in medicine. *Metabolism* 69:S36. <https://doi.org/10.1016/j.metabol.2017.01.011>
- Hernandez-Suarez A, Sanchez-Perez G, Toscano-Medina K, Perez-Meana H, Portillo-Portillo J, Sanchez V, García Villalba LJ (2019) Using twitter data to monitor natural disaster social dynamics: a recurrent neural network approach with word embeddings and

- kernel density estimation. *Sensors* 19(7):1746. <https://doi.org/10.3390/s19071746>
- Holshue ML, DeBolt C, Lindquist S, Lofy KH, Wiesman J, Bruce H, Spitters C, Ericson K, Wilkerson S, Tural A et al (2020) First case of 2019 novel coronavirus in the united states. *New England Journal of Medicine*. <https://doi.org/10.1056/nejmoa2001191>
- Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X et al (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet* 395(10223):497. [https://doi.org/10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5)
- Hung M, Lauren E, Hon ES, Birmingham WC, Xu J, Su S, Hon SD, Park J, Dang P, Lipsky MS (2020) Social network analysis of covid-19 sentiments: Application of artificial intelligence. *J Med Internet Res* 22(8):e22590. <https://doi.org/10.2196/22590>
- Iqbal M, Al-Obeidat F, Maqbool F, Razzaq S, Anwar S, Tubaishat A, Khan MS, Shah B (2021) Covid-19 patient count prediction using lstm. *IEEE Transactions on Computational Social Systems* 8(4):974
- Khatua A, Khatua A, Cambria E (2019) A tale of two epidemics: Contextual word2vec for classifying twitter streams during outbreaks. *Information Processing & Management* 56(1):247. <https://doi.org/10.1016/j.ipm.2018.10.010>
- Kim SC, Hawkins KH (2020) The psychology of social media communication in influencing prevention intentions during the 2019 us measles outbreak. *Computers in Human Behavior* 111:106428
- Kruspe A, Häberle M, Kuhn I, Zhu XX (2020) Cross-language sentiment analysis of european twitter messages during the covid-19 pandemic. arXiv preprint [arXiv:2008.12172](https://arxiv.org/abs/2008.12172)
- Li L, Gao L, Zhou J, Ma Z, Choy D, Hall D (2021) Can social media data be utilized to enhance early warning: Retrospective analysis of the us covid-19 pandemic. medRxiv
- Lipton ZC, Berkowitz J, Elkan C (2015) A critical review of recurrent neural networks for sequence learning. arXiv preprint [arXiv:1506.00019](https://arxiv.org/abs/1506.00019)
- Medina CP, Ramon MRR (2003) Using TF-IDF to Determine Word Relevance in Document Queries. *Proceedings of the first instructional conference on machine learning*, Piscataway, NJ USA pp. 133–142. <https://doi.org/10.15804/tner.2015.42.4.03>
- Mengistie TT, Kumar D (2021) in *2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC)* (IEEE, 2021), pp. 444–449. <https://doi.org/10.1109/ICAIIIC51459.2021.9415191>.
- Mukhtar S (2020) Pakistanis' mental health during the covid-19. *Asian Journal of Psychiatry* 51:102127. <https://doi.org/10.1016/j.ajp.2020.102127>
- Murthy D, Powell AB, Tinati R, Anstead N, Carr L, Halford SJ, Weal M (2016) Bots and political influence: A sociotechnical investigation of social network capital. *International journal of communication* 10:20
- Oliphant T (2006) NumPy. <https://github.com/numpy/numpy/>. [Online; accessed 19-July-2020]
- Oyeyemi SO, Gabarron E, Wynn R (2014) Ebola, twitter, and misinformation: a dangerous combination?. *Bmj* 349. <https://doi.org/10.1136/bmj.g6178>.
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E (2011) Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12:2825
- Pérez-Escoda A, Jiménez-Narros C, Perlado-Lamo-de Espinosa M, Pedrero-Esteban LM (2020) Social networks' engagement during the covid-19 pandemic in spain health media vs. healthcare professionals. *Int J Environ Res Public Health* 17(14), 5261
- Powers DM (2020) Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. arXiv preprint [arXiv:2010.16061](https://arxiv.org/abs/2010.16061)
- Sahni H, Sharma H et al (2020) Role of social media during the covid-19 pandemic: Beneficial, destructive, or reconstructive? *International Journal of Academic Medicine* 6(2):70
- Sepp Hochreiter TUM, Jurgen Schmidhuber I (1997) Long Short-Term Memory. *Neural Computation* 9(8):1735
- Team GB (2015) TensorFlow. <https://github.com/tensorflow/tensorflow/>. [Online; accessed 19-July-2020]
- Twitter scraper (2018) <https://github.com/taspinar/twitterscraper/>. [accessed Aug. 05, 2018]
- Violos J, Tserpes K, Varlamis I, Varvarigou T (2018) Text classification using the n-gram graph representation model over high frequency data streams. *Frontiers in Applied Mathematics and Statistics* 4:41. <https://doi.org/10.3389/fams.2018.00041>
- Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, Wang B, Xiang H, Cheng Z, Xiong Y et al (2020) Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. *Jama* 323(11):1061. <https://doi.org/10.1001/jama.2020.1585>
- Worldometers. <https://www.worldometers.info/coronavirus/> (2004). [Online; accessed 27-October-2021]
- Yang Y, Su Y (2020) Public voice via social media: Role in cooperative governance during public health emergency. *International Journal of Environmental Research and Public Health* 17(18):6840
- Zhang Y, Lyu H, Liu Y, Zhang X, Wang Y, Luo J (2020) Monitoring depression trend on twitter during the covid-19 pandemic. arXiv preprint [arXiv:2007.00228](https://arxiv.org/abs/2007.00228)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.