Multimedia Tools and Applications
https://doi.org/10.1007/s11042-023-17744-1

1236: EXPLAINABLE ARTIFICIAL INTELLIGENCE SOLUTIONS FOR IN-THE-WILD HUMAN BEHAVIOR ANALYSIS

Convolutional neural network and ensemble machine learning model for optimizing performance of emotion recognition in wild

Nazik Alturki[1] · Muhammad Umer[2] · Amal Alshardan[1] · Oumaima Saidani[1] · Andrea F. Abate[3] · Imran Ashraf[4]

Received: 16 September 2023 / Revised: 30 October 2023 / Accepted: 23 November 2023

Abstract

Emotions play a pivotal role in our everyday interactions, serving as a crucial indicator of a speaker's influence. Among various means of expression, facial emotions hold a special prominence compared to hand gestures, body movements, and more. The expressions on the face serve as a primary canvas for conveying these emotions. Numerous studies have introduced approaches and models for recognizing emotions. However, it is important to note that many of these approaches have been predominantly tested on datasets gathered in controlled environments. In contrast, real-world environments are dynamic and unpredictable, introducing numerous challenges to emotion recognition. The current study pursues a two-way approach to emotion recognition. Firstly, the research introduces a novel face mask dataset, which is designed for emotion recognition in a real-time setting. This dataset is utilized in conjunction with a convolutional neural network (CNN), and multiple image processing techniques are applied to the manually labeled face mask dataset. The second phase explores emotion recognition through textual data, employing a range of features, including hand-crafted, Word Embedding, Fast Text embedding, and Transformer models. The study compares the models' performance using original features versus convoluted features. The study aims to provide valuable insights into emotion recognition through both textual and face-masked data, contributing to our understanding of this important field.

Keywords In-the-wild emotion recognition · Facemask · Deep learning · Image processing · Region of interest detection

Abbreviations

Acronyms	Definition
AI	Artificial Intelligence

✉ Imran Ashraf
imranashraf@ynu.ac.kr

Extended author information available on the last page of the article

ANN	Artificial Neural Network
CNN	Convolutional Neural Network
COVID-2019	CoronaVirus disease 2019
ETC	Extra Tree Classifier
LR	Logistic Regression
RF	Random Forest
RILFD	Real Image-based Labelled Face Mask Dataset
SGD	Stochastic Gradient Descent
SVM	Support Vector Machine
TN	True Negative
TP	True Positive
WHO	World Health Organization
VC	Voting Classifier
IDF	Inverse Document Frequency
TF	Term Frequency
BERT	Bidirectional Encoder Representations from Transformers
NLP	Natural Language Processing
BiERU	Bidirectional Emotional Recurrent Unit

1 Introduction

The emergence of new infectious diseases requires a suitable definition of cases which in turn are significant for the surveillance of health services and clinical diagnosis. For the establishment of effective interventions and speed of spread, tracking the cases over a time period is crucial. Coronavirus was declared a pandemic on the 11th of March, 2019 by the World Health Organization (WHO) [1]. The outbreak of COVID-19 has caused havoc around the globe, influencing the societal and financial aspects of countries. It is one of the biggest pandemics faced by the whole world. The life of every human being is greatly affected by the COVID-19 pandemic [2]. The global pandemic puts a lot of pressure on the medical and financial situation of every country. In 2020, WHO claimed that almost 75% of the world countries' healthcare systems were affected by this pandemic. In the beginning, the spread of COVID-19 was limited to some Asian and European countries, but in the second half of 2020, COVID-19 spread to 220 countries of the world. The most distressing aspect of this virus is that the latest iteration of COVID-19 spreads at an exponential rate and eventually takes the forefront as the primary factor behind fatalities in numerous countries. The death toll of this deadly virus has reached millions [3].

Now the world has changed completely and we are living in a situation of never-ending change, in which how we relate and talk with others has been altered continuously [4]. Sharing data progression about the pandemic in public is challenging regarding control, reaction, readiness, government improvements, and media broadcasting [4, 5]. The emergency due to COVID-19 is making certain circumstances that are uncommon for communities relating to health [6]. In the advanced world, communication and information-related technologies are significantly developed [7]. Likewise, COVID-19-related information and comments are continuously spreading on social media. It creates tension and anxiety in the general public and corrupts the information provided by the health authorities [8]. Lockdown urges people to share and read COVID-19 experiences across various social networking sites. Social media platforms with plenty of information either fake or real led people to spend most of the time

[4]. People express their opinions about the COVID-19 pandemic, its medication, and how different countries are dealing with this global pandemic on social networking sites. A large amount of data is generated regarding COVID-19 across a range of social networking sites; hence, the need for automatic approaches to analyze and classify the sentiments of the textual content posted by a user is very important to obtain insights into people's sentiments and plan accordingly [9].

The surge in negative propaganda through social media platforms is becoming a critical issue globally. These comments are creating great trouble in public. This situation causes a significant increase in online data, but false and negative information is affecting the mindset of people. Such misinformation can mislead the measures conveyed by the government and authorities [10]. False information related to COVID-19 can create panic in public and can lead to alarming situations, and authorities need to confirm original and genuine information [11]. In literature, some researchers worked on public opinion on the news concerning the COVID-19 epidemic [12–14]. Twitter, a microblogging and social media platform where people share messages known as 'tweets' has become the leading platform for sharing such information. About five hundred million tweets and two hundred billion tweets per annum have been posted on Twitter which makes it a significant data conversation platform among the public globally [15]. Chakraborty et al. explained that a large number of COVID-19 tweets contain positive sentiments. However, users primarily concentrate on amplifying tweets with offensive language in the word frequency of their posts [16].

Authors applied dual contrastive learning approach for the detection of epistemic emotion [17]. Zhang et al. [18] proposed a self-training classifier for emotion recognition using EEG signals. Nie et al. [19] designed a framework based on common sense graph knowledge and extracted topics from dialogues. Another study explored emotion detection in short text using a multilabel KNN model [20]. False and fake information on social networking sites like Twitter twists peoples' minds and people criticize governmental policies. Therefore, analyzing the sentiments of people regarding COVID-19 is very helpful in determining overall sentiments taking corresponding actions, and devising appropriate policies for corrective actions. In this regard, machine learning models can play an important role in analyzing massive data which requires several experts. Various methods of feature engineering, both independently and in combination with multiple text preprocessing techniques, are employed to extract essential features. These features play a pivotal role in facilitating the precise classification of COVID-19 tweets into two distinct categories: positive and negative. The following contributions are made by this work, which employs machine learning to examine the opinions of participants about COVID-19.

Emotion recognition in real-world, uncontrolled environments has become increasingly significant across a range of fields. This capability plays a pivotal role in human-computer interaction and has far-reaching effects in areas such as affective computing, healthcare, marketing, and social robotics [21]. The capacity to understand and appropriately respond to human emotions is particularly valuable in applications like virtual assistants and educational software, as it fosters more authentic and empathetic interactions. Furthermore, in the realm of healthcare, the skill to detect emotional states can significantly contribute to mental health assessments and therapeutic interventions. It's important to note that emotion recognition in natural settings isn't confined to tech-focused sectors; it extends to areas like security and law enforcement, where the identification of emotional signals can be instrumental in threat detection. With the application of machine learning methods and access to copious data resources, this field is continuously progressing, delivering profound insights into human behavior that have diverse and invaluable applications.

Detection of face masks serves various practical purposes in real-world scenarios, including remote monitoring of individuals and real-time biometrics, emotion recognition, and many others. Notably, individuals with malicious intent often conceal their faces by covering the mouth area, presenting a challenge known as occulted face detection. Prior research has addressed this issue by leveraging techniques such as head and shoulder shape analysis [22]. However, the complexity escalates when multiple individuals need remote monitoring for compliance with face mask regulations. Additionally, the global outbreak of the novel coronavirus disease (COVID-19) has necessitated widespread mask usage, along with other precautionary measures recommended by the World Health Organization (WHO) to combat its rapid spread and the lack of a specific cure [23, 24]. The use of face masks has emerged as a crucial strategy to mitigate the dire consequences of COVID-19. It has garnered widespread acceptance as an effective measure in curbing the virus's transmission. Consequently, this has placed global pressure on individuals to adhere to social distancing and embrace protective measures to prevent the virus's contagious spread. Despite significant vaccination efforts, COVID-19 and its variants continue to pose threats, emphasizing the ongoing importance of consistent mask usage in containing its dissemination and safeguarding individuals from potential exposure.

- This research work is based on emotion recognition in two ways a) through textual data and b) through face-masked recognition.
- The first phase of this research work that is based on textual data emotion recognition makes use of many types of features such as hand-crafted, Word Embedding, Fast Text embedding, and Transformer models.
- In the first phase, this study examined the performance of various supervised machine learning models including extra tree classifier (ETC), random forest (RF), Naive Bayes (NB), gradient boosting machine (GBM), stochastic gradient descent (SGD), logistic regression (LR), and a voting classifier that merges SGD and LR. Additionally, the research evaluated different feature engineering techniques such as term frequency (TF), term frequency-inverse document frequency (TF-IDF), and the fusion of both feature generation methods (TF+ TF-IDF) and FastText embedding. Furthermore, the study compared the performance of the models using the original features versus convoluted features.
- The efficacy of the proposed approach is contrasted with the performance of existing models. Furthermore, bidirectional encoder representation from transformers (BERT) models, such as BERT, PHS-BERT, and BioALBERT, were employed to compare their performance.
- In the second phase of experiments, this research work makes use of its own designed face mask dataset with a convolutional neural network (CNN) for emotion recognition in a real-time environment.
- Multiple-image processing techniques are applied to the manually labeled face mask dataset.

The paper's structure is as follows: In Section 2, an examination of pertinent research and methodology is presented. Section 3 delineates the dataset, the associated preparation processes, and the intricate details of the suggested approaches. Furthermore, this section offers insights into the cutting-edge models incorporated in the research. The results and observations are presented in Section 4. Finally, Section 6 concludes the study by summarizing the findings and suggesting future research directions.

2 Related work

Face mask-based emotion recognition has become an increasingly relevant and challenging task in the wake of the COVID-19 pandemic, where widespread mask-wearing has become a norm. Emotion recognition is vital in various applications, including healthcare, human-computer interaction, and security systems. However, the presence of face masks poses a significant obstacle to traditional facial expression analysis techniques, which rely on the visibility of key facial landmarks. In recent research, innovative solutions have emerged to address this challenge. One approach involves leveraging the visible facial regions above the mask, such as the eyes and eyebrows, to infer emotions accurately. These regions contain valuable information, including changes in eye gaze, eyelid movements, and eyebrow configurations, which can indicate different emotional states [25].

Recent studies have explored the development of deep learning-based models to perform emotion recognition in masked faces. Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have been applied to learn discriminative features from the limited facial information available. Transfer learning techniques have also been employed, using pre-trained models on unmasked datasets to boost performance on masked face images. Additionally, the incorporation of thermal imaging and infrared technologies has shown promise in capturing subtle changes in skin temperature around the eyes, which can provide additional cues for emotion recognition [26]. Despite these advances, ongoing research aims to improve the accuracy and robustness of mask-based emotion recognition systems, recognizing the continued importance of this technology in the post-pandemic era. Authors proposed a model framework based on relational reason in [27]. Another study used feature fusion for human detection in crowd [28].

Although numerous studies have explored emotion recognition in unobstructed facial expressions or partially obscured faces, there is a limited body of research that specifically examines how face masks affect the accuracy of emotion recognition. Some indications suggest that the presence of a mask may impede the ability to perceive facial emotions [29]. Nevertheless, these tests were conducted on relatively small sample sizes, and additional research, including a control group to assess the specific impacts of masks, is essential. A study performed emotion recognition on occluded faces and also investigated threat n masked faces [30]. Another study examined a more extensive and varied participant pool, and it also explored how masks influenced the perceived intensity of expressions in individuals exhibiting traits associated with autism [31]. The research examined how the use of face masks affected communication within healthcare settings [32].

The authors in [33] conducted two tasks involving adults who were either wearing surgical face masks or not. These tasks aimed to evaluate language processing skills and the ability of young children to recognize emotions. The study's results revealed that while younger children exhibited lower accuracy compared to older children, the presence of face masks did not substantially hinder their fundamental language processing skills. However, it had a notable impact on the children's accuracy in recognizing emotions, making masked angry faces more readily distinguishable and masked happy and sad faces less distinguishable. The children's age and their social-emotional skills also played a role in these findings. Deep learning models have been employed by researchers for emotion recognition in masked faces [34].

Numerous artificial intelligence (AI)-based sentiment analysis techniques have been extensively utilized across various domains. Amidst the COVID-19 pandemic, sentiment analysis has been conducted on various topics. Textual analysis finds extensive utility in a variety of applications, including the identification of sarcasm and irony [35], the mining of sentiment and opinions [36], the detection of disinformation [37], as well as the extraction of medical-related information from texts, among other uses. Twitter data stands prominently as a primary source for emotional assessment and maintains a leading role in collecting health-related feedback. In India, news sentiment analysis is done on 24000 COVID-19 tweets [38]. Another research study looks into how COVID-19 affects people psychologically and how it affects behavior [39]. It claims that individuals are in an intense situation and with a high level of depression due to the COVID-19 news. Various classifiers have been applied to short-text information. NB and LR give average results on the short text and do not do well with extensive texts [40].

Xue et al. [41] worked on 4 million COVID-19 tweets with 25 distinct hashtags from March 1, 2020, to April 21, 2020. Each of the 13 topics has five classes. Latent Dirichlet allocation (LDA) is utilized for the recognition of uni-gram, bi-gram, and themes, and for analyzing sentiments in the tweets. While dealing with health-related issues, results and accuracy are vital. Another study detected emotions using 2500 short-text advertisements, and 2500 lengthy-text messages [42]. Depression has been identified as the most dominant emotion. The main cause of the depression is due to the long stay at home, fear of being positive for COVID-19, and joblessness [43]. To study emotions, bidirectional encoder representations from Transformers (BERT) are also utilized. BERT can assign single, as well as, multi-labels [44]. A notable aspect of the model is its capacity to consider emojis, which serve as a valuable means of expressing emotions. Drias et al. [45] worked on the FP-growth algorithm-based sentiment analysis.

Imamah and Rachman [46] worked on the Twitter sentiment analysis. They consider the tweets specifically about COVID-19. The dataset used in the study is the dataset containing COVID-19-related tweets. The authors collected 355384 tweets related to COVID-19 and used the term weighting TF-IDF and LR methods to classify the tweets. The highest accuracy of 94.71% is achieved. Chintalapudi et al. [47] worked on COVID-19 tweets' sentiment analysis. The research study consists of different classifiers. The study also includes data analysis using BERT, and a comparison of the performance of BERT and other models also analyzed like SVM, LR, and long short-term memory (LSTM).

Lopez et al. [48] conducted a study centered on the COVID-19 discussions taking place on Twitter and their correlation with government interventions. They undertook a multilingual analysis of Twitter data from diverse user groups and nations, aiming to scrutinize prominent policy responses as well.

Naseem and colleagues [49] conducted a study using a benchmark dataset of tweets related to COVID-19. They analyzed a dataset consisting of 90,000 COVID-19-related tweets spanning from February 2020 to March 2020. Various machine learning models, such as Support Vector Machine (SVM), Random Forest (RF), Naive Bayes (NB), and Decision Trees (DT), were deployed to categorize the tweets into positive, negative, and neutral classes. To establish a baseline for the machine learning classifier, feature extraction methods like TF-IDF and FastText embedding were applied. The research also incorporated multiple deep learning models. The findings indicate that BERT and its variants outperform other techniques, including TF-IDF and word embedding.

The range of sentiment analysis approaches employed in earlier studies lacks in several aspects. First, the focus is mostly on improving the models regarding the architecture, hyper-parameters fine-tuning, etc., and the feature extraction part is under-explored. Secondly, often complex models are tried which have high computational complexity. Thirdly, the effectiveness of BERT models is not extensively investigated. Therefore, this study performs sentiment analysis with several machine learning classifiers in combination with various feature extraction approaches like TF, TF-IDF, FastText, and a combination of TF and IDF. In addition, BERT models are also used for performance comparison.

3 Materials and methodologies

3.1 Phase 1: Emotion recognition using facemask

In this section, we delve into the description of the proposed framework, the process of dataset acquisition, and the sequential steps undertaken within the framework. The schematic representation of the proposed framework's workflow is visualized in Fig. 3. The initial phase involves the collection of a dataset encompassing instances of both masked and unmasked classes. Subsequently, the collected images undergo a comprehensive four-stage image preprocessing pipeline. This series of image processing stages is vital due to the inherent challenges posed by variations in skin tone and lighting conditions, rendering face mask detection a complex task. The primary objective of these image processing steps is to standardize the input images, thus enhancing their suitability for subsequent analysis. Following this preprocessing phase, a specialized deep learning model is employed to ascertain the presence or absence of a face mask. Additionally, two pre-trained deep learning models, which have been fine-tuned for this specific task, are also integrated into the framework to bolster its performance.

3.1.1 Real image-based labeled face mask dataset

The availability of a meticulously labeled and well-structured face mask dataset is a valuable asset for the research community. While there are existing datasets for experimentation in this domain, they come with certain limitations. These limitations encompass factors such as small image sizes and the utilization of internet-sourced images. Furthermore, some datasets incorporate simulated face masks onto subjects' faces. In response to these limitations, this study introduces a novel labeled dataset known as the "Real Image-based Labeled Face Mask Dataset" (RILFD). The RILFD dataset is thoughtfully curated by capturing photographs of individuals with face masks using a Nikon D5300 digital single-lens reflex (DSLR) camera, equipped with an 18/140 lens. The images within this dataset possess dimensions of 11903 pixels in width and 13096 pixels in height[1].

In total, the RILFD dataset comprises 750 images, each featuring 250 unique individuals. The data collection process involves initially photographing individuals with masks happy and subsequently capturing images of the same individuals without facemasks unhappy. Each image is meticulously annotated with corresponding labels, indicating the presence of emotions ('happy') or ('happy') in face masks. The inclusion of these high-resolution images, along with their associated labels, facilitates a detailed and nuanced analysis of the task of face mask detection. The primary objective of the RILFD dataset is to provide a meticulously labeled dataset that is well-suited for training machine learning models. Notably, several example images from this dataset are showcased in Fig. 1.

[1] https://github.com/MUmerSabir/FaceMaskDataset

Fig. 1 RILFD dataset examples with both labels. Figure 2(a) represents the facemask unhappy and Fig. 2(b) shows the facemask happy

3.1.2 Image preprocessing

In the context of face mask detection, a crucial preliminary step involves image preprocessing, which comprises four distinct stages within the proposed framework. The framework is structured in a cascading manner, encompassing both image processing and a deep Convolutional Neural Network (CNN) for the effective detection of face masks. The primary objective of the image preprocessing stages is to standardize and enhance the input facial images, ultimately facilitating more efficient training of deep learning models. These four image processing stages are detailed below:

Stage 1: In the initial stage, the original data undergoes a filtering process. Specifically, a filter with dimensions ([0, -1, 0], [-1, 6, -1], [0, -1, 0]) is applied to the images, as depicted in Fig. 2(b).

Stage 2: Advancing to the third stage, the images are transformed from the Blue, Green, and Red (BGR) color space to the YUV color space. This transformation retains the luminance (Y) component at full resolution while reducing the resolutions of the U and V color components. This reduction in U and V resolutions not only simplifies the training model but also acknowledges the greater significance of luminance in image information. The BGR to YUV conversion process is illustrated in Fig. 2(c).

Stage 3: In the penultimate stage, the images are normalized by reverting back to the BGR color space. This step serves to smoothen the images and applies histogram normalization, as visually represented in Fig. 2(d).

Stage 4: The final stage encompasses contrast conversion and image resizing. Here, the dimensions of the input image are reduced, as the original dataset images are of substantial size at 11903x13096 pixels. This resizing operation reduces the image dimensions to $120 \times 120 \times 3$, as demonstrated in Fig. 2(e). This resizing operation contributes to a reduction in computational time during the subsequent detection process.

3.1.3 Proposed convolutional neural network

Face mask detection models have witnessed significant advancements in recent years, particularly gaining momentum during the COVID-19 pandemic, as highlighted by Asif et al. (2021) [50]. Notably, the field of face detection has experienced substantial enhancements through the introduction of deep learning models, such as Convolutional Neural Networks (CNNs). Researchers have harnessed the power of deep CNN models in various domains, ranging from numeric data analysis [51] to text data analysis [52]. Specifically, the realm of face detection

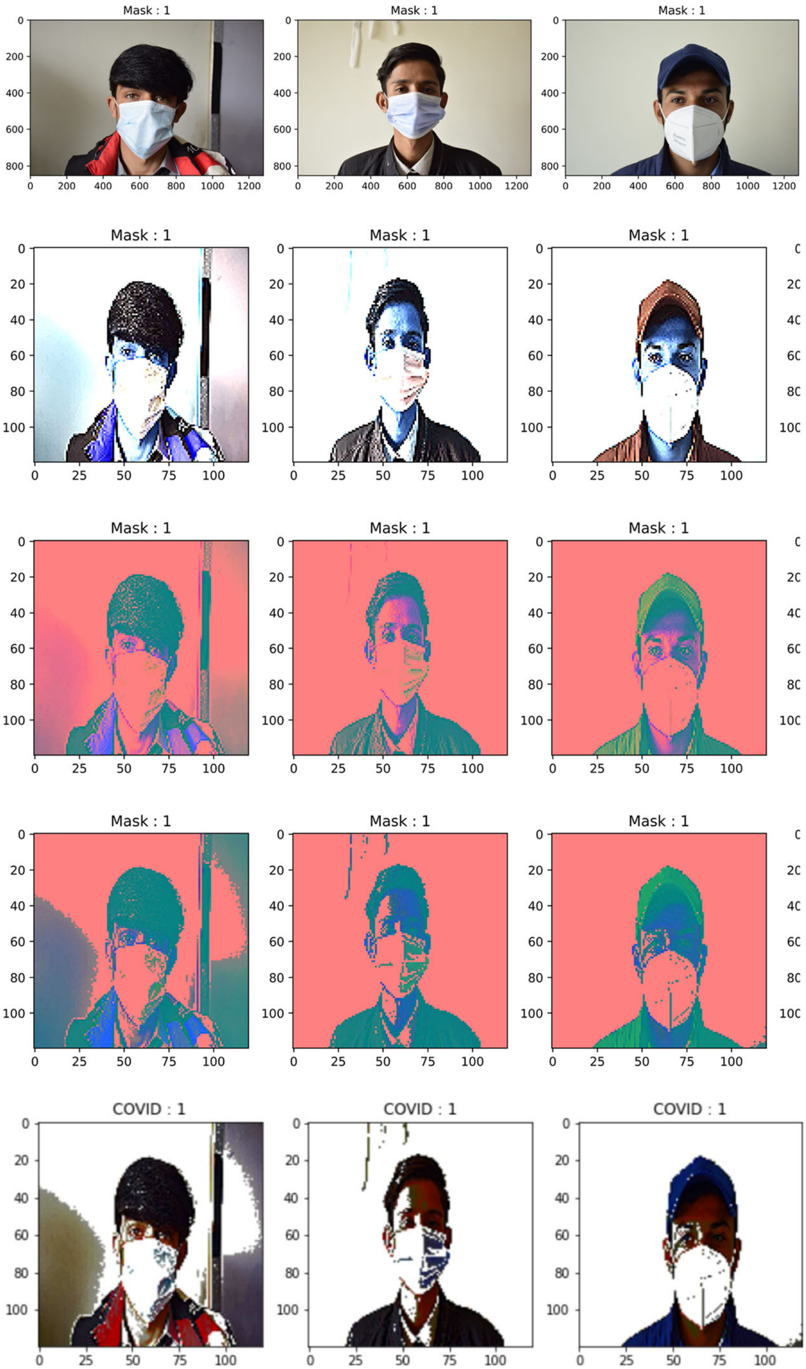

Fig. 2 Image preprocessing steps followed in the experiments, **(a)** Original Image, **(b)** Kernel is applied for edge detection, **(c)** BGR image is converted to YUV to get Y_0, and **(d)** Histogram Equalization of YUV image, and **(e)** YUV image is converted to BGR image

predominantly relies on CNN-based techniques, as elucidated in the study by Mohan et al. (2021) [53], which underscores their exceptional performance in effectively addressing a multitude of challenges, including variations in pose, low-resolution image inputs, varying illumination conditions, and diverse forms of noise. Although the field of face mask detection has encountered different types of occlusions, it remains a relatively underexplored research area. Within this context, our paper leverages a deep learning model based on CNNs, a well-established and widely employed approach renowned for its effectiveness in various object detection tasks.

The Convolutional Neural Network (CNN) plays a pivotal role in the realm of computer vision tasks due to its remarkable ability to extract features efficiently, all while maintaining a relatively low computational cost, as underscored in Chauhan et al.'s work [54]. One of its highly advantageous applications is binary image classification. CNN achieves this through the convolution operation, wherein it convolves feature maps or images with various kernels to extract high-level features. However, a recurring challenge lies in the quest to design an improved CNN architecture. Among the well-recognized CNN models, the Inception model stands out, as discussed in Szegedy et al.'s work [55]. Additionally, the field introduced the MobileNet model, designed for object detection on mobile devices, emphasizing a low computational cost, as proposed by Wang et al. [56]. This model strategically leverages depth and channel-wise convolutions to reduce computational demands effectively.

A CNN typically comprises convolutional layers, pooling layers, and fully connected layers, each serving distinct functions. Within the CNN model, a kernel or filter, represented as a numerical matrix, convolves over the image, transforming it based on the filter's values. After each convolution step, the image size is progressively reduced, facilitating efficient processing within a limited timeframe.

3.2 Phase 2: Emotion recognition through textual data

This section presents the approaches and procedures employed in the experiments. It also includes a comprehensive discussion of the dataset description, preparation procedures, machine learning models, and performance assessment metrics utilized in the experiments.

3.2.1 Dataset

The dataset including COVID-19-related tweets was collected through the IEEE DataPort [57]. It has attributes of 'tweet_id' and 'sentiment_score' of tweets. Various hashtags and keywords are used for tweet derivation to ensure that the text discusses the COVID-19-related situation. Tweet_ID provided by the IEEE platform is hydrated to obtain whole tweets-related data. There are 11858 records in the dataset. Keywords that are used to acquire data are $covid - 19$, $pandemic$, $quarantine$, $sarscov2$, $n95$, and $lockdown$ etc. To get a detailed view of tweets from the COVID-19 dataset, a visual presentation is used. Firstly, the most used terms from the tweet sentiment dataset are given. Several terms are used most commonly like coronavirus', 'COVID-19', and 'lockdown'.

3.2.2 Data preprocessing

A dataset comprises samples or observations, where each sample has several attributes or features that correlate to the basic characteristics of data objects. These qualities or properties are commonly denoted as data dimensions or aspects of the dataset [58]. The dataset is obtained from various sources and may contain redundant, incomplete, or irrelevant data that needs to be removed. Data preprocessing involves organizing and cleaning the dataset to eliminate irrelevant content and make it suitable for training machine learning models. The dataset employed in this investigation is partially organized in a non-standard manner, includes irrelevant information, and was acquired from the IEEE data repository. In order to minimize the influence of extraneous data on the predictive method, text preprocessing is indispensable. Employing extensive datasets can lead to extended training times, and incorporating stop words may diminish predictive precision. Therefore, essential preprocessing measures like stemming, converting uppercase letters to lowercase, punctuation addition, and the removal of non-essential words that don't contribute substantive meaning to the text are imperative.

3.2.3 Techniques for extracting features

To train classifiers, textual data needs to be represented as vectors. This necessitates the transformation of textual data into numerical form while retaining all original information. Various techniques can be employed for this purpose, such as TF and FastText embedding, which assigns a vector value to each word. Another frequently utilized approach is the bag of words (BoW). Nevertheless, BoW has its limitations, particularly in the case of short-character tweets, which diminish its effectiveness. Furthermore, the accuracy of BoW-based methods is constrained by the sparse occurrence of words in text comments or tweets [59]. In our study, we opted for the TF method for this transformation [60]. TF converts a collection of text documents into an integer matrix, with each value representing the frequency of each word's occurrence in the document, capturing the word frequency within each one.

In the proposed approach [61], we have utilized TF, TF-IDF, and a combination of both to represent features. The TF-IDF method assigns lower weights to common terms that appear in most of the texts, and higher weights to words that only exist in a portion of the documents. This technique penalizes frequently occurring terms by giving them lower weights while giving higher weights to uncommon words present in a particular document.

3.2.4 Machine learning models

This research employs various machine learning models including probability-based models, ensemble learning classifiers, and regression-based models to classify tweets. The models were implemented using Scikit learn [62, 63] and Python. The study also incorporates meta-algorithms that combine multiple machine learning approaches into a single predictive model to reduce variance (bagging), and bias (boosting), and improve predictions [64]. Specifically, this work focuses on the use of ensemble learning-based models for sentiment analysis of tweets related to COVID-19. Here's a summary of the machine learning models used in this study.

3.2.5 Random forest

RF is a machine learning algorithm that combines multiple decision trees to improve efficiency and reduce overfitting of the model [65]. This approach operates through the creation of decision tree classifiers using various input data samples, followed by the aggregation of the outcomes from each tree classifier to serve as an ensemble learner. The algorithm produces multiple decision trees, attributing the output class for classification or forecasting the average value for regression at each node within the tree. Random Forest is a classification technique that assembles numerous decision trees for data analysis, rendering it a widely embraced machine learning algorithm, appreciated for its straightforwardness and versatility. The algorithm can deliver good results without the need for adjusting hyperparameters.

3.2.6 Gradient boosting machine

GBM stands as an ensemble machine learning algorithm that relies on a loss function for constructing an optimal additive model [66]. Its operation is iterative, where it employs the loss function in each step to minimize the error rate. The primary goal of GBM is to determine the expected results of the target variable for the subsequent model to minimize prediction errors. The target variable's outcome is contingent on substantial alterations in predictions that influence the overall error. If there is a significant reduction in the error rate, it results in a higher value for the subsequent target prediction, which, in turn, diminishes the prediction error as the predictions from the next model approach closer to the target variable.

3.2.7 Extra tree classifier

ETC represents a tree-based ensemble machine learning model with a functioning akin to the RF model [67]. It is often referred to as a random tree because it constructs trees based on the actual data samples without the utilization of bootstrap data. ETC chooses the root node from randomly selected samples, employing the Gini index. The algorithm was developed with the intention of creating trees while considering the incorporation of numerical inputs and determining the optimal cut-point to diminish variance at each node and reduce computational complexity. This model has demonstrated its reliability in addressing intricate and high-dimensional problems. In contrast to RF, which generates continuous piece-wise approximations, ETC produces multi-linear approximations.

3.2.8 Logistic regression

LR is a machine learning algorithm specialized in handling classification tasks, operating under a probability-driven approach [68]. It employs a logistic function to model binary variables, revealing the association between the dependent and independent variables via the Sigmoid function. LR quantifies the correlation coefficient, serving as a metric to gauge the connection between the target variable and the independent variables. LR is widely employed for binary classification endeavors, aimed at forecasting the likelihood of an instance belonging to a specific class. The correlation coefficient assesses the relationship between values within a range from -1 to 1, illustrating the congruence between actual values and expected values.

3.2.9 Naive bayes

NB is a powerful algorithm that works on the 'Bayes' theorem [69]. It operates by calculating the likelihood of data instances belonging to a specific class and conditional probability. The highest probability class is considered the final prediction. It assumes that a specific feature is unrelated to any other feature from the data. If its assumption proves true on the data, then it shows good results even on a small-sized training dataset.

3.2.10 Stochastic gradient descent

The SGD classifier is a machine learning model focused on discovering the optimal parameter for establishing a connection between predicting and actual outcomes [70]. It enhances the objective function while exhibiting traits of smoothness. This model exhibits accelerated learning when confronted with extensive datasets, surpassing the performance of gradient descent. Furthermore, it attains convergence swiftly by creating a batch from the dataset, which is then utilized to determine the gradient at each iteration step.

3.2.11 Voting classifier

A voting classifier or ensemble model is referred to as a collection of several models combined to make aggregated predictions [71, 72]. It works by averaging the output of each machine learning classifier in the ensemble and thus based on the majority of voting it then predicts the output. The voting classifier that is utilized in this study leverages the benefits of LR and SGD. It works by averaging the predicted results obtained from these two models. The target value with the highest probability is selected as the output of the model. It combines the benefits of the models that have been integrated and produces efficient results.

3.3 Convolutional neural networks for textual emotion recognition

The CNN model used in this study comprises a max-pooling layer, an embedding layer, a flatten layer, and a 1D convolutional layer. The embedding layer uses 100 features from the COVID-19 tweets dataset with a vocabulary size of 10000, 100 output dimensions, and a 25 input length. Following the embedding layer, there is a 1D convolutional layer that employs 500 filters, a ReLU activation function, and a 2 kernel size. After the 1D convolutional layer, a max-pooling layer is applied with a pool size of 2 to identify the key characteristics of the output. Finally, a flattened layer is used to convert the output into a 1-dimensional array, as machine learning models typically perform better on 1-dimensional data [73].

Suppose that the COVID-19 dataset X is a tuple set fs_i, tc_i), where fs is the feature set,tc is the target class column, and I is the index of the tuple; then transform the training set in the embedding layer into the needed format of input as follows

$$LE = embedding\_layer(Vs, Os, I) \tag{1}$$

$$GOs = LE(f_s) \tag{2}$$

where GO_s shows the embedding layer output that can develop the convolutional layer inputs, and the embedding layer comprises three different parameters; three input lengths, two output dimensions, and one vocabulary size Vs.

The input size to the model is directly related to the vocabulary size. this study uses the fixed vocabulary size of 10000, which indicates that the model can accept inputs with sizes ranging from 0 to 10000. The output dimension parameter GOs has a value of 100 since the data's output dimension after passing through the embedding layer is 100. The input length I, which displays the 100 features in the COVID-19 dataset, is the third most crucial parameter. The input layer processes the data. The output of this input data is produced for further processing by the CNN model. The embedding layer's input dimensions are $GOs = (None, 500, 100)$.

$$1D - Convs = CNN(F, Qs, ET) \leftarrow GOs \tag{3}$$

where 1D convolutional layers output is represented by the 1D-Convs.

The output of the 1D convolutional layer is taken from the embedding layer. In this study, we used the 500 filters, i.e., $F = 500$, in the CNN layer and the kernel size is $Qs = 2$. ReLU has employed the activation function, which sets the negative values in the $1D - Convs$ output matrix to zero while leaving the other unchanged.

$$f(x) = max(0, E)s \tag{4}$$

The max pooling layer maps CNN's major characteristics. A 2 pool is utilized for the feature map. $Amap$ denotes the features gained after pooling, $Ts = 2$ denotes the pooling window size, and $S - 2$ denotes the stride:

$$Zf = Amap = \lfloor (1 - T_s)/S \rfloor + 1 \tag{5}$$

To convert the 2D data into a 1D flattened layer is used. The reason for using the 1D data is that the machine learning models work effectively with the 1D data. These 25,00 characteristics can be used to train machine learning models.

3.3.1 Proposed framework for textual emotion recognition

The objective of this study is to create an accurate model for classifying COVID-19 sentiments using appropriate features. To accomplish this goal, the research utilizes machine learning models to conduct sentiment analysis on tweets associated with COVID-19. The workflow of the methodology used in this study is shown in Fig. 3. After preparing the data, the sentiment analysis task is performed using the dataset of COVID-19-related tweets obtained from the IEEE data repository. The study analyzes the impact of various feature extraction techniques by conducting experiments with different feature extraction techniques. During the training phase of the learning algorithms, the influence of different feature engineering methods, including TF, TF-IDF, TF+TF-IDF, and FastText embedding, as well as complex features, is taken into account. The complete workflow is explained in Algorithm 1. Starting with the data acquisition, tweets are labeled using TextBlob. It is followed by the preprocessing steps to remove noise and unnecessary and redundant information before the data can be used as a feature. TF-IDF is utilized for feature engineering.

$$TF(t) = \frac{N}{D} \tag{6}$$

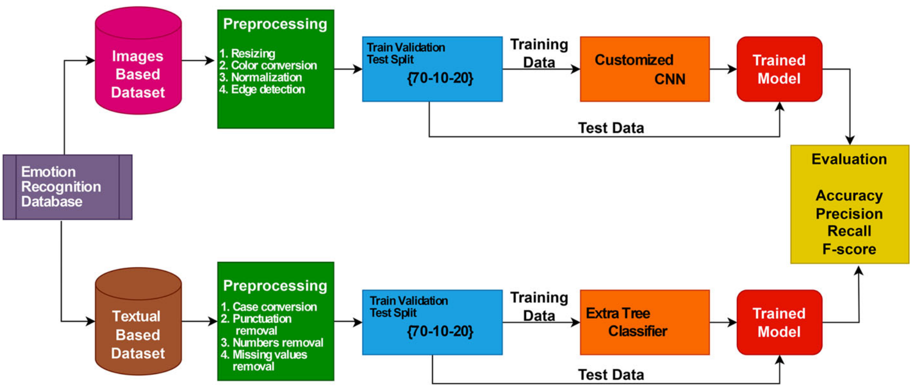

Fig. 3 Architecture of the proposed approach for emotion recognition in the wild environment

$$IDF(t) = \log \frac{d}{dt} \tag{7}$$

TF-IDF can be calculated as

$$W_{t,d} = TF_{t,d}\left(\frac{N}{D_{f,t}}\right) \tag{8}$$

Afterward, the data is divided into training and testing sets, and the ETC model is supplied with the extracted features for training. The trained model is evaluated using hidden samples which provide the sentiment labels in the form of 'positive' or 'negative' sentiments.

Each feature generation technique results are obtained using supervised machine learning models. Machine learning models utilized in this study include RF, ETC, LR, GBM, NB, SGD, and VC (LR+SGD). The dataset has a training-validation-testing ratio of 0.7, 0.1, to 0.2, respectively. In the end, several well-known performance assessment criteria are used to examine the models' performance.

Algorithm 1 Workflow process algorithm.

Input: COVID-19 Tweets.
Output: COVID-19 Tweets sentiment.
Steps Involved:
Step 1: Extraction of COVID-19 Tweets
Step 2: Data Annotation Using TextBlob.
Step 3: Data Preprocessing
Step 4: Extract the Features (TF, TF-IDF, and FastText) and convoluted features using (7), (8), and (3).
Step 5: Data Splitting into Training (70%), Validation (10%), and Testing (20%) set.
Step 6: Train Extra Tree Classifier using training data
Step 7: Classify and predict using the trained model ((Positive) (Negative)).
Step 8: Compute the average values of evaluation parameters on predicted tweets.
Step 9: Output: COVID-19 tweets sentiments

The performance assessment matrices employed in this study are accuracy, precision, recall, and F1 score. These matrices are built based on four terms; true positive (TP), true negativity (TN), false positive (FP), and false negative (FN). TP denotes positive examples that have been successfully categorized, negative occurrences that TN has correctly classified, misclassified positive instances are depicted by FP, and misclassified FN denotes negative instances. We may assess the Accuracy, Precision, Recall, and F-score based on these criteria.

The frequently used measures for assessing classifier performance are accuracy, precision, recall, and F-score. They are calculated as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

$$Precision = \frac{TP}{TP + FP} \tag{10}$$

$$Recall = \frac{TP}{TP + FN} \tag{11}$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{12}$$

4 Results and discussion

4.1 Phase 1: Results of face-masked emotion recognition

The second phase of experiments involves employing the aforementioned image preprocessing steps. The experimental outcomes are summarized in Table 1, showcasing great performance with image preprocessing. Notably, all learning models exhibit substantially enhanced performance when image preprocessing is applied. Among these models, Proposed CNN delivers superior results, boasting an accuracy of 98.36%, precision of 98%, recall of 99%, and an F1-score of 98% when compared to the machine learning algorithm. The complete results of all classifiers are shown in Fig. 4.

4.2 Phase 2: Results of textual data based emotion recognition

The main objective of this research is to conduct sentiment analysis on tweets related to COVID-19, utilizing machine learning models. The Python programming language's Scikit package is employed to construct these models. To examine the sentiment analysis of COVID-19 tweets, experimental findings from various feature engineering techniques are studied. The supervised machine learning models including RF, NB, GBM, ETC, LR, SGD, and VC (LR+SGD) are compared, using TF, TF-IDF, and FastText Embedding techniques to obtain results.

4.3 Evaluating classifiers utilizing TF features

As depicted in Table 2, provides a contrast among machine learning models utilizing the TF method, with regard to accuracy, precision, recall, and F1 score. The ETC model shows superior performance compared to other TF models, achieving a 95.60% accuracy rate. Similarly, the SGD model also demonstrates significant results with TF, achieving a 94.97% accuracy rate in classifying tweets related to COVID-19. Both the SGD and ETC models exhibit a 95% value for precision, recall, and F1 score. NB and GBM, on the other hand, performed poorly in the sentiment analysis of COVID-19 tweets. GBM shows an 87.30% accuracy, 87% recall, 89% precision, and 86% F1 score. NB showed an 88.77% accuracy, 89% recall, 89% precision, and 90% F1 score.

Table 1 Face mask emotion detection using deep learning models on RILFD dataset with image preprocessing techniques

Models	Accuracy	Precision	Recall	F1-score
RF	95.41%	96%	95%	95%
GBM	89.61%	91%	92%	91%
ETC	93.28%	94%	94%	94%
NB	89.89%	91%	92%	91%
LR	94.17%	93%	90%	92%
SGD	91.28%	90%	89%	89%
VC(LR+SGD)	96.55%	97%	96%	96%
Customized CNN	98.36%	98%	99%	98%

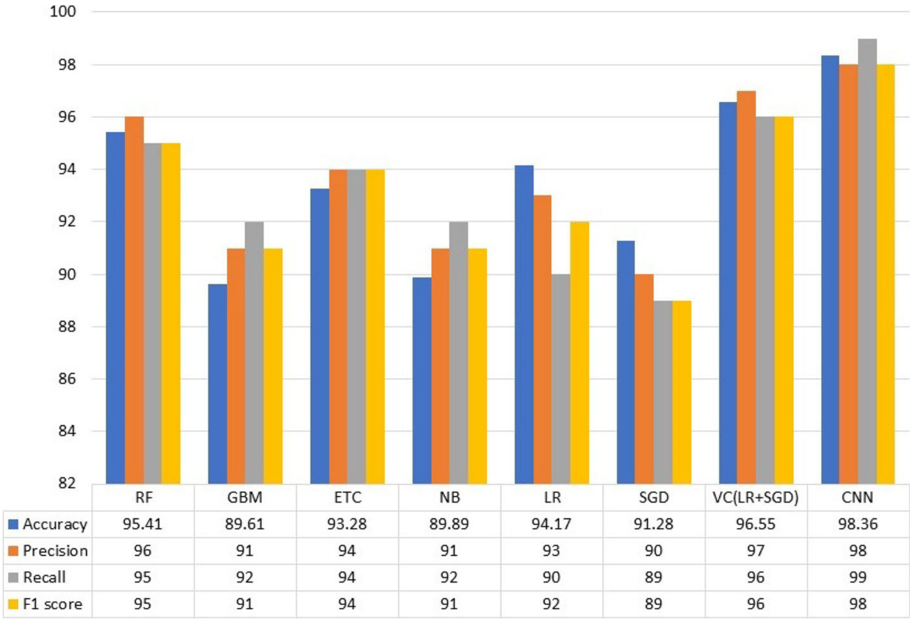

Fig. 4 Comparison of classifiers for face-mask-based emotion recognition detection

4.4 Performance of machine learning classifiers using TF-IDF

Table 3 displays the outcomes concerning the categorization of COVID-19-associated tweets by supervised machine learning algorithms employing TF-IDF attributes. Upon examining the findings, it is evident that the application of TF-IDF has led to enhancements in the performance of both the SGD and ETC models. SGD showed a 95.10% accuracy using TF-IDF in classifying tweets which is the second-highest accuracy. NB and GBM do not perform better than other machine learning models using TF-IDF, but their performance is still improved as compared to the results obtained using TF features. ETC shows the highest precision of 96%, SGD shows 95%, and RF shows 93% precision scores. The lowest value of precision is achieved by GBM which is 89%. The highest value of recall which is 96%, is achieved by the ETC model and SGD attains the second highest value with 95% recall, and RF stood third with a 93% recall score. The highest F1 score of 96% is also achieved by the ETC model.

Table 2 Comparison of machine learning classifiers using TF features

Model	Accuracy	Precision	Recall	F-score
RF	93.25%	94%	94%	93%
GBM	87.30%	89%	87%	86%
ETC	95.60%	95%	95%	95%
NB	88.77%	89%	89%	88%
LR	93.94%	94%	93%	93%
SGD	94.97%	95%	95%	95%
VC(LR+SGD)	93.14%	94%	93%	93%

Table 3 Assessing classifiers utilizing TF-IDF

Model	Accuracy	Precision	Recall	F-score
RF	92.89%	93%	93%	93%
GBM	87.35%	88%	88%	87%
ETC	95.47%	96%	96%	96%
NB	88.39%	90%	90%	89%
LR	90.19%	92%	91%	90%
SGD	95.10%	95%	95%	95%
VC(LR+SGD)	91.20%	92%	91%	91%

4.5 Comparison of classifiers using FastText embedding

The efficiency of supervised machine learning models is also compared by utilizing FastText embedding for COVID-19 tweets. FastText embedding has proven to be an effective text categorization technique [74]. Experimental results in Table 4 show that the supervised machine learning models do not show robust results when used with FastText. ETC obtains the best performance with an 89.46% accuracy; however, it is lower than what was achieved using TF and TF-IDF features, i.e., 95.60% and 95.47%, respectively. The experimental results reveal that the utilization of the FastText embedding for feature representation does not exert any impact on the performance of any classifier. Notably, the ETC classifier attains the highest F1 score at 89% when employing FastText embedding, albeit it is 8% lower than the F1 score achieved with the use of TF-IDF.

4.6 Machine learning models and feature fusion

In order to substantiate the effectiveness, efficiency, and resilience of machine learning models, we conducted experiments that involved feature fusion (TF+TF-IDF). While assessing the sentiment of COVID-19-related tweets, it becomes evident that LR, SGD, and VC (LR+SGD) surpass other models, achieving a 93% rating in terms of accuracy, precision, recall, and F1 score when employing feature fusion (TF+TF-IDF), as demonstrated in Table 5. ETC also outperforms alternative models, achieving a 93% rating in precision, recall, and F1 score. Notably, NB and GBM exhibit suboptimal performance when using feature fusion, with results akin to those attained using TF-IDF and TF. Specifically, NB demonstrates enhanced performance with an 89.93% accuracy through feature fusion compared to FastText embedding, which achieves 69.49% accuracy. Similarly, GBM exhibits a marginally improved

Table 4 Comparison of classifiers using FastText embedding

Model	Accuracy	Precision	Recall	F-score
RF	88.33%	89%	88%	88%
GBM	85.54%	86%	85%	85%
ETC	89.46%	91%	90%	89%
NB	69.49%	73%	70%	71%
LR	83.12%	83%	83%	83%
SGD	84.63%	84%	84%	84%
VC(LR+SG)	83.14%	83%	83%	83%

Table 5 Comparison of classifiers using feature fusion (TF+TF-IDF.)

Model	Accuracy	Precision	Recall	F-score
RF	91.38%	92%	92%	91%
GBM	86.94%	87%	86%	85%
ETC	92.39%	91%	91%	91%
NB	89.93%	89%	89%	89%
LR	93.70%	93%	93%	93%
SGD	93.81%	93%	93%	93%
VC(LR+SGD)	93.01%	93%	93%	93%

performance with feature fusion in comparison to FastText embedding when applied to sentiment analysis of COVID-19-related tweets.

4.7 Discussions

Based on the above-mentioned findings, it is evident that machine learning models have been successful in distilling public sentiment from the dataset consisting of tweets. Various feature engineering methods, encompassing TF, TF-IDF, TF+TF-IDF, and FastText embedding, have been employed across multiple investigations. Among these models, ETC, RF, LR, SGD, and VC (LR+ SGD) have all exhibited accuracy exceeding 93% when TF was applied, considering precision, recall, and F1 score.

In the case of supervised learning models utilized in these experiments, improved performance was observed when TF-IDF was employed in contrast to TF, with the exception of GBM. Conversely, FastText embedding exhibited suboptimal performance when utilized within supervised machine learning models for sentiment analysis of tweets. The outcomes strongly suggest that FastText embedding does not enhance the classifiers' efficiency in tweet categorization. Furthermore, to substantiate the effectiveness of supervised machine learning models, an evaluation was conducted using feature fusion involving both TF and TF-IDF. LR, SGD, and VC (LR+SGD) performed notably well with feature fusion. As seen in Table 5, LR and SGD achieved commendable results, reaching 93% across accuracy, recall, precision, and F1 score. Notably, the results were closely aligned with those achieved by their ensemble counterpart, VC (LR+SGD). According to the experimental findings, the ETC classifier outshone all other models when employing TF, TF-IDF, TF+TF-IDF, and FastText embedding. It's noteworthy that ETC demonstrated the ability to categorize sentiments in tweets with remarkable precision, yielding a 95% accuracy alongside 96% in terms of precision, recall, and F1 score, particularly when using TF-IDF features.

4.8 Performance of machine learning models using CNN features

A second series of experiments is conducted utilizing CNN features to assess the effectiveness of machine learning and the suggested model to increase the measurement accuracy for the categorization of COVID-19 tweets. Experimental findings are displayed in Table 6. The proposed tree-based model ETC performs better than the other learning models, with the greatest accuracy of 99.99%, according to the findings of the machine learning models employing the CNN features. It shows a 6.0% increase in ETC's performance when compared to its performance with the TF+TF-IDF features and the original features.

Table 6 Comparison of classifiers using CNN features

Model	Accuracy	Precision	Recall	F-score
RF	99.35%	99%	98%	98%
GBM	97.96%	97%	96%	96%
ETC	99.99%	99%	99%	99%
NB	84.93%	86%	90%	88%
LR	88.72%	91%	89%	90%
SGD	90.19%	88%	92%	90%
VC(LR+SGD)	92.93%	96%	95%	95%

Expanding the feature set is the goal of employing CNN model features, which is anticipated to improve the accuracy of linear models. This is done by employing CNN-extracted features to train and evaluate machine learning models. The accuracy of models considerably increases when they are employed with convoluted features from CNN when CNN is used for feature extraction, which is due to the rise in the number of features. The performance of machine learning models using CNN features is shown in Fig. 5.

5 Comparison with encoding-decoding models

The Transformer model, most notably exemplified by models like BERT (Bidirectional Encoder Representations from Transformers), is a powerful deep learning architecture designed for natural language processing tasks. Unlike earlier models that used recurrent or convolutional networks, Transformers rely on a self-attention mechanism to process input

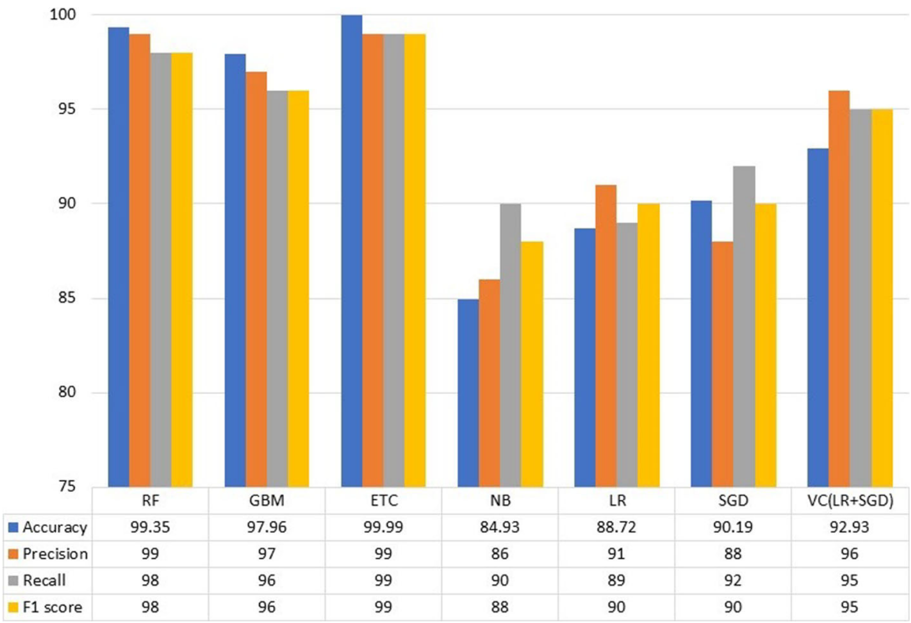

Fig. 5 Comparison of classifiers using CNN features

data. This mechanism allows the model to weigh the importance of different words in a sentence concerning each other, which is highly effective for understanding context in language.

To contrast the proposed strategy with the most recent model, this study applied BERT, PHS-BERT, and BioALBERT models. Using a sizable dataset of 3.3 billion words, the success of BERT has been analyzed. BERT is trained on the 800 million words BooksCorpus of Google and the 2.5 billion word Wikipedia. BERT is unique from other conventional models since it can read simultaneously in both directions. Bidirectionality is the term for this capacity, which was made possible with the help of transformers. BERT is outperforming in different 11 NLP tasks. PHS-BERT is proposed for tasks related to public health surveillance and has proved its robustness [75]. BioALBERT 1.1 is a large model that is trained on biomedical corpus and outperforms other models on many datasets [76].

Results of the transformer-based models are given in Table 7. It can be observed that all three models have achieved an accuracy greater than 90%. BioALBERT has attained the highest accuracy with 95.07% which is slightly higher than the proposed ETC model with TF-IDF features. If we compare both models in terms of complexity, it can be noticed that BioALBERT is a multilayer complex deep learning model, and trained on millions of parameters that require high computational power. Contrarily, ETC is a simple machine-learning model and can be executed on lower-power machines. Thus ETC is more suitable for performing COVID-19 tweet analysis.

6 Conclusion

Emotions are an important part of our daily conversation and show the influence of a speaker. While various modes of expression exist, facial emotions stand out as the primary canvas for conveying these emotions. Although numerous studies have introduced approaches and models for recognizing emotions, many of these have been primarily tested in controlled environments, neglecting the complexity of real-world scenarios. To address this gap, the present study adopts a multifaceted approach to emotion recognition. It initiates by introducing a face mask dataset, explicitly designed for real-time emotion recognition. This dataset is a valuable contribution to the field, as it allows for the examination of emotions in a dynamic and unpredictable environment. By integrating a convolutional neural network (CNN) and employing various image processing techniques, this study adds a practical dimension to the understanding of emotion recognition by achieving 98.36% accuracy. The second phase of experiments explores textual data, employing a diverse range of features, from handcrafted attributes to advanced techniques like Word Embedding, Fast Text embedding, and Transformer models. The comparative analysis of models using both original and convoluted features sheds light on their performance in varying contexts. The proposed ETC using the CNN feature outperforms other models with 99.99% accuracy. This research not only highlights the importance of considering real-world challenges in emotion recognition but

Table 7 Comparison with transformer-based models

Model	Accuracy	Precision	Recall	F-score
BERT	90.61%	89%	91%	90%
PHS-BERT	92.88%	90%	93%	92%
BioALBERT	95.07%	95%	95%	95%
Proposed	94.74%	95%	95%	95%

also provides a comprehensive framework for addressing these challenges. The combination of textual and face-masked data analysis adds depth and relevance to the field, offering insights that can inform applications across a wide range of domains, from healthcare to human-computer interaction and beyond. The findings presented in this paper contribute to the growing body of knowledge on emotion recognition, making strides toward more robust and practical solutions for understanding and interpreting emotions in diverse contexts.

Acknowledgements This research was funded by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R333), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Funding This research was funded by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R333), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Data Availability The datasets generated during and/or analyzed during the current study is available from the corresponding author on reasonable request.

References

1. Bai Y, Yao L, Wei T, Tian F, Jin DY, Chen L, Wang M (2020) Presumed asymptomatic carrier transmission of COVID-19. Jama 323(14):1406
2. Lades LK, Laffan K, Daly M, Delaney L (2020) Daily emotional well-being during the COVID-19 pandemic. Br J Health Psychol 25(4):902
3. Depoux A, Martin S, Karafillakis E, Preet R, Wilder-Smith A, Larson H (2021) Covid-19 coronavirus / death toll. https://www.worldometers.info/coronavirus/coronavirus-death-toll/
4. Depoux A, Martin S, Karafillakis E, Preet R, Wilder-Smith A, Larson H (2021) Covid-19 coronavirus / death toll. https://www.worldometers.info/coronavirus/coronavirus-death-toll/
5. Staszkiewicz P, Chomiak-Orsa I, Staszkiewicz I (2020) Dynamics of the COVID-19 contagion and mortality: Country factors, social media, and market response evidence from a global panel analysis. IEEE Access 8:106009. https://doi.org/10.1109/ACCESS.2020.2999614
6. Guo YR, Cao QD, Hong ZS, Tan YY, Chen SD, Jin HJ, Tan KS, Wang DY, Yan Y (2020) The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak - an update on the status. Mil Med Res 7(1):1
7. Mittal M, Battineni G, Goyal LM, Chhetri B, Oberoi SV, Chintalapudi N, Amenta F (2020) Cloud-based framework to mitigate the impact of COVID-19 on seafarers' mental health. Int Marit Health 71(3):213
8. Garcia LP, Duarte E (2020) Infodemic: excess quantity to the detriment of quality of information about covid-19
9. 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
10. Apuke OD, Omar B (2021) Fake news and COVID-19: modelling the predictors of fake news sharing among social media users. Telemat Inform 56:101475
11. Al-Zaman M et al (2021) COVID-19-related social media fake news in India. Journalism and Media 2(1):100
12. Depoux A, Martin S, Karafillakis E, Preet R, Wilder-Smith A, Larson H (2020) The pandemic of social media panic travels faster than the covid-19 outbreak
13. Gao J, Zheng P, Jia Y, Chen H, Mao Y, Chen S, Wang Y, Fu H, Dai J (2020) Mental health problems and social media exposure during COVID-19 outbreak. Plos One 15(4):e0231924
14. Ahmad AR, Murad HR (2020) The impact of social media on panic during the COVID-19 pandemic in Iraqi Kurdistan: online questionnaire study. J Med Internet Res 22(5):e19556
15. Stats IL (2017) Retrieved July 22(2017):2017
16. Chakraborty K, Bhatia S, Bhattacharyya S, Platos J, Bag R, Hassanien AE (2020) Sentiment analysis of COVID-19 tweets by deep learning classifiers–a study to show how popularity is affecting accuracy in social media. Appl Soft Comput 97:106754
17. Liu Z, Wen C, Su Z, Liu S, Sun J, Kong W, Yang Z (2023) Emotion-semantic-aware dual contrastive learning for epistemic emotion identification of learner-generated reviews in MOOCs. IEEE Trans Neural Netw Learn Syst

18. Zhang X, Huang D, Li H, Zhang Y, Xia Y, Liu J (2023) Self-training maximum classifier discrepancy for EEG emotion recognition. CAAI Trans Intell Technol
19. Nie W, Bao Y, Zhao Y, Liu A (2023) Long dialogue emotion detection based on commonsense knowledge graph guidance. IEEE Trans Multimed
20. Liu X, Shi T, Zhou G, Liu M, Yin Z, Yin L, Zheng W (2023) Emotion classification for short texts: an improved multi-label method. Humanit Soc Sci Commun 10(1):1
21. Tzirakis P, Zhang J, Schuller BW (2018) In: 2018 IEEE international conference on acoustics, speech and signal processing (ICASSP), IEEE, pp 5089–5093
22. Zhang T, Li J, Jia W, Sun J, Yang H (2018) Fast and robust occluded face detection in ATM surveillance. Pattern Recognit Lett 107:33
23. Leung NH, Chu DK, Shiu EY, Chan KH, McDevitt JJ, Hau BJ, Yen HL, Li Y, Ip DK, Peiris J et al (2020) Respiratory virus shedding in exhaled breath and efficacy of face masks. Nat Med 26(5):676
24. Feng S, Shen C, Xia N, Song W, Fan M, Cowling BJ (2020) Rational use of face masks in the COVID-19 pandemic. Lancet Respir Med 8(5):434
25. Almaev T, Won J (2020) Symmetry 12(10):1665
26. Zhang S, Liu X, Cui J, Yao L, Zhang L, Yang J (2021) IEEE Transactions on Systems, Man, and Cybernetics: Systems pp 1–13
27. Liu Y, Li G, Lin L (2023) IEEE Trans Pattern Anal Mach Intell
28. Zhou X, Zhang L (2022) An effective feature pyramid network for crowded human detection. Appl Intell 52(11):12556
29. Marini M, Ansani A, Paglieri F, Caruana F, Viola M (2021) The impact of facemasks on emotion recognition, trust attribution and reidentification. Sci Rep 11(1):5577
30. Grahlow M, Rupp CI, Derntl B (2022) The impact of face masks on emotion recognition performance and perception of threat. PLoS One 17(2):e0262840
31. Pazhoohi F, Forby L, Kingstone A (2021) Facial masks affect emotion recognition in the general population and individuals with autistic traits. PLoS One 16(9):e0257740
32. Bani M, Russo S, Ardenghi S, Rampoldi G, Wickline V, Nowicki S Jr, Strepparava MG (2021) Behind the mask: Emotion recognition in healthcare students. Med Sci Educ 31(4):1273
33. Bourke L, Lingwood J, Gallagher-Mitchell T, López-Pérez B (2023) The effect of face mask wearing on language processing and emotion recognition in young children. J Exp Child Psychol 226:105580
34. Mukhiddinov M, Djuraev O, Akhmedov F, Mukhamadiyev A, Cho J (2023) Masked face emotion recognition based on facial landmarks and deep learning approaches for visually impaired people. Sensors 23(3):1080
35. Naseem U, Razzak I, Eklund P, Musial K (2020) In: 2020 International Joint Conference on Neural Networks (IJCNN), IEEE, pp 1–7
36. Naseem U, Khan SK, Razzak I, Hameed IA (2019) In: Australasian joint conference on artificial intelligence, Springer, pp 381–392
37. Aggarwal CC, Reddy CK (2013) Data clustering: algorithms and applications. Chap Man And Hall
38. Barkur G, Vibha GBK (2020) Sentiment analysis of nationwide lockdown due to COVID 19 outbreak: Evidence from India. Asian J Psychiatr 51:102089
39. Li S, Wang Y, Xue J, Zhao N, Zhu T (2020) The impact of COVID-19 epidemic declaration on psychological consequences: a study on active Weibo users. Int J Environ Res Public Health 17(6):2032
40. Samuel J, Ali G, Rahman M, Esawi E, Samuel Y et al (2020) Covid-19 public sentiment insights and machine learning for tweets classification. Information 11(6):314
41. Xue J, Chen J, Hu R, Chen C, Zheng C, Zhu T (2020) arXiv:2005.12830
42. Kleinberg B, van der Vegt I, Mozes M (2020) arXiv:2004.04225
43. Li I, Li Y, Li T, Alvarez-Napagao S, Garcia-Gasulla D, Suzumura T (2020) In: International conference on innovative techniques and applications of artificial intelligence Springer, pp 358–370
44. Feng Y, Zhou W (2020) arXiv:2006.08581
45. Drias HH, Drias Y (2020) medRxiv
46. Rachman FH, et al (2020) In: 2020 6th Information technology international seminar (ITIS), IEEE, pp 238–242
47. Chintalapudi N, Battineni G, Amenta F (2021) Sentimental analysis of COVID-19 tweets using deep learning models. Infect Dis Rep 13(2):329
48. Lopez CE, Vasu M, Gallemore C (2020) arXiv:2003.10359
49. Naseem U, Razzak I, Khushi M, Eklund PW, Kim J (2021) COVIDSenti: a large-scale benchmark Twitter data set for COVID-19 sentiment analysis. IEEE Trans Comput Soc Syst
50. Asif S, Wenhui Y, Tao Y, Jinhai S, Amjad K (2021) In: 2021 4th International conference on artificial intelligence and big data (ICAIBD), IEEE, pp 70–75

51. Manzoor M, Umer M, Sadiq S, Ishaq A, Ullah S, Madni HA, Bisogni C (2021) RFCNN: traffic accident severity prediction based on decision level fusion of machine and deep learning mode. IEEE Access 9:128359
52. Karim M, Missen MMS, Umer M, Sadiq S, Mohamed A, Ashraf I (2022) Citation context analysis using combined feature embedding and deep convolutional neural network model. Appl Sci 12(6):3203
53. Mohan P, Paul AJ, Chirania A (2021) In: Innovations in electrical and electronic engineering, Springer, pp 657–670
54. Chauhan R, Ghanshala KK, Joshi R (2018) In: 2018 First international conference on secure cyber computing and communication (ICSCCC) IEEE, pp 278–282
55. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016) In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2818–2826
56. Wang W, Li Y, Zou T, Wang X, You J, Luo Y (2020) Mob Inf Syst 2020
57. Lamsal R (2020) Applied intelligence pp 1–15
58. Bow ST (2002) Pattern recognition and image preprocessing. CRC Press
59. Sriram B, Fuhry D, Demir E, Ferhatosmanoglu H, Demirbas M (2010) In: Proceedings of the 33rd international ACM SIGIR conference on research and development in information retrieval, ACM, pp 841–842
60. learn S (2018) Scikit-learn feature extraction with tf/idf. https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.CountVectorizer.html. Accessed 5 Apr 2019
61. learn S (2018) Scikit-learn feature extraction with tf/idf. https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.Tfidf/. Accessed 5 Apr 2019
62. Hackeling G (2017) Mastering Machine Learning with scikit-learn, Packt Publishing Ltd
63. learn S (2017) Scikit-learn classification and regression models. http://scikitlearn.org/stable/supervised_learning.html. Accessed 10 Apr 2019
64. Araque O, Corcuera-Platas I, Sanchez-Rada JF, Iglesias CA (2017) Enhancing deep learning sentiment analysis with ensemble techniques in social applications. Expert Syst Appl 77:236
65. Breiman L (2001) Machine learning (1)
66. Friedman JH (2001) Annals of statistics pp 1189–1232
67. Sharaff A, Gupta H (2019) In: Advances in computer communication and computational sciences, Springer, pp 189–197
68. Genkin A, Madigan D, Lewis DD (2007) Large-scale Bayesian logistic regression for text categorization. Technometrics 49(3):291
69. Perez A, Larranaga P, Inza I (2006) Supervised classification with conditional Gaussian networks: increasing the structure complexity from naive Bayes. Int J Approx Reason 43(1):1
70. Gardner WA (1984) Learning characteristics of stochastic-gradient-descent algorithms: a general study, analysis, and critique. Signal Process 6(2):113
71. Rupapara V, Rustam F, Shahzad HF, Mehmood A, Ashraf I, Choi GS (2021) Impact of SMOTE on imbalanced text features for toxic comments classification using RVVC model. IEEE Access 9:78621
72. Khalid M, Ashraf I, Mehmood A, Ullah S, Ahmad M, Choi GS (2020) GBSVM: sentiment classification from unstructured reviews using ensemble classifier. Appl Sci 10(8):2788
73. Rustam F, Ishaq A, Munir K, Almutairi M, Aslam N, Ashraf I (2022) Incorporating CNN features for optimizing performance of ensemble classifier for cardiovascular disease prediction. Diagnostics 12(6):1474
74. Stein RA, Jaques PA, Valiati JF (2019) An analysis of hierarchical text classification using word embeddings. Inf Sci 471:216
75. Naseem U, Lee BC, Khushi M, Kim J, Dunn AG 2022) arXiv:2204.04521
76. Naseem U, Khushi M, Reddy V, Rajendran S, Razzak I, Kim J (2021) In: 2021 International joint conference on neural networks (IJCNN), IEEE, pp 1–7

Authors and Affiliations

Nazik Alturki[1] · Muhammad Umer[2] · Amal Alshardan[1] · Oumaima Saidani[1] · Andrea F. Abate[3] · Imran Ashraf[4]

Nazik Alturki
namalturki@pnu.edu.sa

Muhammad Umer
umer.sabir@iub.edu.pk

Amal Alshardan
amalshardan@pnu.edu.sa

Oumaima Saidani
ocsaidani@pnu.edu.sa

Andrea F. Abate
abate@unisa.it

1 Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia

2 Department of Computer Science, The Islamia University of Bahawalpur, 63100 Bahawalpur, Pakistan

3 Department of Computer Science, University of Salerno, 84084 Fisciano, Italy

4 Information and Communication Engineering, Yeungnam University, 38541 Gyeongsan, Korea