Int. J. Appl. Math. Comput. Sci., 2022, Vol. 32, No. 2, 323–334
DOI: 10.34768/amcs-2022-0024

HYBRID DEEP LEARNING MODEL–BASED PREDICTION OF IMAGES RELATED TO CYBERBULLYING

MAHMOUD ELMEZAIN [a,b,*], AMER MALKI [b], IBRAHIM GAD [a], EL-SAYED ATLAM [a,b]

[a]Faculty of Science
Tanta University
Tanta 31527, Egypt

[b]College of Computer Science and Engineering
Taibah University
Yanbu 966144, Saudi Arabia
e-mail: {mahmoud_osman_s,satlam}@yahoo.com

Cyberbullying has become more widespread as a result of the common use of social media, particularly among teenagers and young people. A lack of studies on the types of advice and support available to victims of bullying has a negative impact on individuals and society. This work proposes a hybrid model based on transformer models in conjunction with a support vector machine (SVM) to classify our own data set images. First, seven different convolutional neural network architectures are employed to decide which is best in terms of results. Second, feature extraction is performed using four top models, namely, ResNet50, EfficientNetB0, MobileNet and Xception architectures. In addition, each architecture extracts the same number of features as the number of images in the data set, and these features are concatenated. Finally, the features are optimized and then provided as input to the SVM classifier. The accuracy rate of the proposed merged models with the SVM classifier achieved 96.05%. Furthermore, the classification precision of the proposed merged model is 99% in the bullying class and 93% in the non-bullying class. According to these results, bullying has a negative impact on students' academic performance. The results help stakeholders to take necessary measures against bullies and increase the community's awareness of this phenomenon.

Keywords: cyberbullying, ResNet50, MobileNetV2, support vector machine.

1. Introduction

Bullying is a behavioral issue that has attracted the attention of educators and policy makers all over the world. It is also defined as a situation in which one or more students repeatedly perpetrate purposeful verbal or physical abuse on one or more peers. Bullying is a serious problem which can affect school achievements, but it is often disregarded (Delprato *et al.*, 2017; Longobardi *et al.*, 2022). Cyberbullying is becoming an increasingly widespread problem all over the world. In essence, cyberbullying is very similar to the form of bullying that many children have unhappily been accustomed to at their schools. The only difference is that it happens online, rather than in person.

Cyberbullying is the act of bullying someone using digital media technology. With the advent of the Internet and mobile devices, a new euphemism for bullying has developed. It is a crime perpetrated by minors who have access to technologies like cell phones, instant messaging, e-mail, chat rooms, or social networking sites like Facebook and Twitter to harass, threaten, or intimidate someone online. Cyberbullying may have serious consequences, especially for children and teenagers which causes a lot of emotional and psychological problems. Anxiety, fear, depression, a low self-esteem and a lack of success at school can be experienced by children who are subjected to such activities (Hellsten *et al.*, 2021; Devidas *et al.*, 2021; Blachnik, 2019).

Social media have over 330 million active users and most of them are in the age ranging from 18 to 65. Children under 18 years old, most of them growing in developed countries, also actively use social media

*Corresponding author

tools for bullying purposes. Due to the amount of personal data shared and interactions among users, social networking sites have offered a platform for cyberbullying by youngsters (Alim, 2015; Elmezain *et al.*, 2021). The age group containing victims of cyberbullying is for teenagers between 12 and 18. Children, young adults (particularly girls), and students are the most prevalent targets of cyberbullies. According to a survey by Sophia (Abaido, 2019) on 7,200 US teenagers in 2014, 59% of teenagers interviewed used social media containing images, text, and videos. There are many examples of bullying in social media, such as harassment, sexting, cyberstalking, deceit, impersonation, and sending harsh messages via email.

According to cyberbullying data from 2019, about 43% of youngsters (mostly girls) have experienced some form of online harassment. More than half of the Internet users in the United States (53%) have experienced cyberbullying, with 37% reporting serious online harassment, such as physical threats (22%), sexual harassment (18%), stalking (18%), and persistent harassment (17%) (Setiana and Besar, 2021).

Bullies use a variety of techniques to agitate individuals with their nasty thoughts. This paper focuses in some cyberbullying, which can be done by flaming, harassing, exclusion, stalking, framing, dissing and trolling (McGuire and Norman, 2018; Cortis and Handschuh, 2015; Ibrahem *et al.*, 2020).

A primary issue with cyberbullying is persistence because digital technologies let young people connect instantaneously and constantly 24 hours a day; finding a respite from cyberbullying can be difficult. If not reported and erased, the vast majority of information given electronically becomes permanent and public. A bad on-line reputation, especially for bullies, may affect college admissions, employment, and other parts of life. Another issue is that cyberbullying is hard to notice. It is harder to detect since teachers and parents may not hear or see it.

The motivation for this paper is the need for more research on different sorts of information linked to cyberbullying, notably advice and help associated with countering it. This information can help victims of cyberbullying of all ages, including many teenagers, find a support network. In this study, about 1200 images associated with cyberbullying were extracted and analyzed. A hybrid model has been developed to classify the images in this data set and make predictions. The features of the data are extracted using ResNet50, EfficientNetB0, Mobilenet, and Xception architectures, and then used as input to a convolutional neural network (CNN) in conjunction with an SVM for prediction after the optimization process (Zohair *et al.*, 2021).

The main points of this study are as follows:

- Bullying has a negative impact on academic performance, and students' social emotional skills can help them deal with it.
- Governments must enforce binding laws for criminalizing the phenomenon of bullying.
- Social emotional skills should be considered in anti-bullying programs.
- Raising awareness of the problem among parents and the society will help to addressing it.
- To the best of our knowledge, this study is the first examination of many diverse types of cyberbullying behaviors. The results support the importance of understanding and preventing exposure to harmful cyberbullying activities among youths.
- The accuracy rate of the proposed merged models with the SVM classifier achieved 96.05%.

The paper is organized as follows: Section 2 presents the related work. Section 3 introduces the proposed methodology. Sections 4 present the experimental observations. In Section 5, the conclusions and proposed future studies are presented.

2. Related work

Bullying is not only a widespread problem, but it is also extremely costly, especially when both victims and perpetrators of bullying suffer from negative consequences throughout their lives, as Sarzosa and Urzúa (2015) point out. Repeating this conduct several times might reveal emotional vulnerability and a great level of psychic pain in the oppressor. According to `StopBullying.gov`, 160,000 children miss school every day in the United States due to a fear of being bullied (representing 15% of all students missing classes); bullying victims are 2 to 9 times more likely to consider suicide than non-bullying victims; and at least 50% of young people's suicides in the United Kingdom are related to these experiences.

There are studies employing deep learning methods, according to the literature review on the issue. Rui Zhao and Chen (2019) employed a data set with five classes. To classify photos, they employed a multi-label classification strategy that included a CNN and an RNN. They mentioned that they employed the CNN to extract features and then the RNN structures to further process these features. In this investigation, the maximum accuracy rate was found to be 92.63%.

Kibriya *et al.* (2015) conducted a survey of 7323, Ghanaian eight graders in 2011 to investigate school bullying. Bullying has a negative influence on arithmetic scores, according to the findings, with the size of the effect being greater in girls. In the case of students who have a female teacher, the impact of bullying is reduced. To validate their findings, the authors used propensity score matching and a set of robustness tests. According to them, anti-bullying policies must take gender into consideration.

Bullying is linked to poor academic achievements in the majority of studies (Le *et al.*, 2005; Kosciw *et al.*, 2012; Ponzo, 2013). According to Boulton and Underwood (1992), bullying victims are more likely to experience melancholy and loneliness at school, as well as have fewer close friends.

According to Fekkes (2006), children who have been bullied are more likely to develop new psychosomatic and psychosocial problems than children who have not been bullied, implying a difficult time dealing with loneliness, anxiety, and depression, which can be linked to academic performance due to the expected struggles students may face when confronted with such challenges.

Previous studies have found that bullying victims are more prone than non-bullied youngsters to develop new psychosomatic and psychosocial issues. These studies, on the other hand, are missing several potential elements that could help us tackle these issues (Malki *et al.*, 2020; Atlam *et al.*, 2022).

3. Methodology

In the following subsection many CNN architectures models such as EfficientNetB0, MobileNetV2, ResNet-50 and Xception in addition to SVM classification (CNN-SVM) models are discussed in detail (Fig. 2).

3.1. Pre-trained models. The study uses CNN models, which are becoming increasingly popular as technology advances (Emine Cengil, 2019; Hark and Turgut Özal, 2019). Three stages are carried out for this work. For the first stage, pre-processing is performed on our own data set, which contains 1200 images of positive and negative bullying.

In the second stage, seven most known pre-trained architectures like MobileNetV2, ResNet50, EfficientNetB0,Xception InceptionResNetV2 and NASNetLarge are used (Fig. 1).

The Mobilenet is the initial model used in this research that was developed by Harjoseputro *et al.* (2020). Instead of using the traditional convolution approach, the depthwise separable convolutions technique is employed to extract features in this model. This technique's feature extraction can be done with 8 or 9 times fewer parameters than a normal convolution procedure. Various improvements were made to the model to make it faster and more efficient (Sandler *et al.*, 2018). Using 1×1 convolutions, the size of feature maps was also lowered.

ResNet-50 is a very deep CNN architecture that uses residual learning to solve the degradation problem. The ResNet-50 architecture was used to fit the training data without transfer learning because the initial weights were derived on a completely different type of data and are therefore unlikely to be of any benefit (Tripathi *et al.*, 2020). ResNet's main features are residual learning and identity mapping. The skip connection acts as an identity mapping for the block inputs. Going deeper in typical convolutional networks has a significant impact on the network's performance. Due to consecutive multiplication, vanishing gradients may occur in these networks, causing the gradient value to become too small during backpropagation. The remaining blocks ensure that reverse propagation is unaffected. To put it another way, residual learning is accomplished by feeding data into the output of a single-layer or multi-layer convolutional network, integrating them, and then applying the ReLU activation function. The output convolution network is given the necessary padding to make it the same size as the input. The network not only solves the issue of vanishing gradients, but it also stimulates feature reuse, which increases the outputs' feature variance (Radhika *et al.*, 2020).

The EfficientNetB0 model is the second one used in this study. The EfficientNet model was created by the GoogleBrain team to improve the 'CNN's performance. According to them, the term "depth" refers to the addition of extra layers between or above the previously existing deep convolution model. As a result, as more computational power and resources are required, the growth in depth layers is considered to increase the performance of our method. The depth, width, and resolution characteristics were all considered when the EfficientNet model was created (Mingxing Tan, 2019).

Xception is a depth-wise separable convolutions-based deep convolution neural network architecture. The "extreme" version of an inception module is known as Xception. The inception module performs numerous transformations on the same input before merging the results, allowing the model to pick which features to adopt and by how much. It is still computationally inefficient. These convolutions occur not just in terms of space, but also in terms of depth as a result, with every additional filter. The researchers working on the Inception module were able to concatenate several layer modifications in parallel as a result of this reduction, resulting in a CNN that was wide and deep (Cleetus *et al.*, 2021).

InceptionResNetV2 is a convolutional neural network based on the Inception family of designs, but with residual connections (which replace the filter concatenation step in the Inception architecture) (Zhang *et al.*, 2021).

The NASNet-Large convolutional neural network is trained on a larger number of photos from a database. The network can categorize photos into item categories and outputs a NASNet-Large convolutional neural network

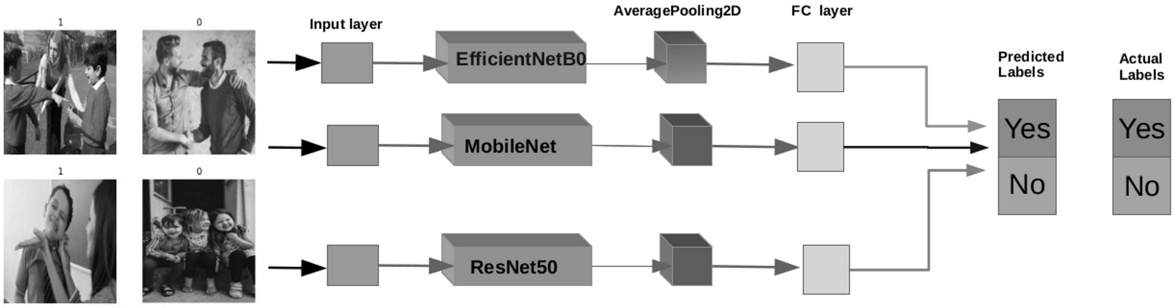

Fig. 1. Estimating cyberbullying with the CNN model.

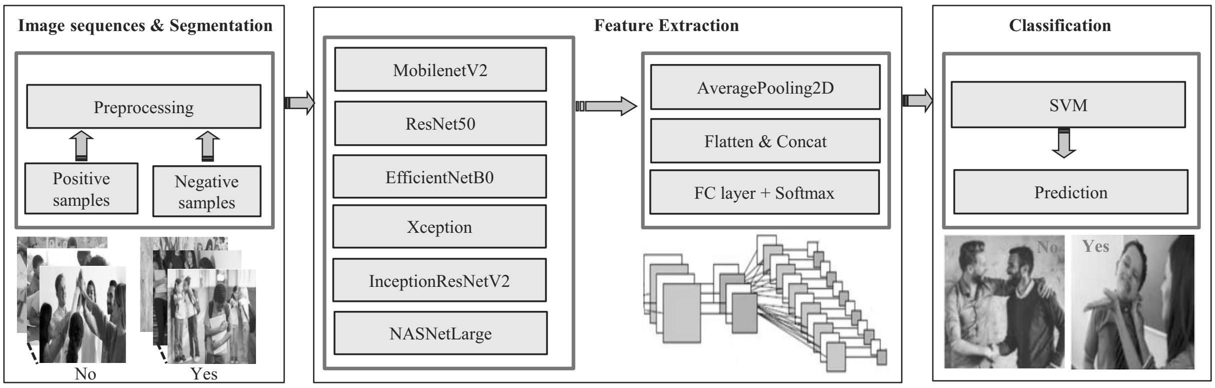

Fig. 2. Estimating cyberbullying with the proposed model.

that has been pre-trained (Zhang *et al.*, 2021).

Figure 1 shows the working logic of the pre-trained EfficientNetB0, MobileNetV2, and ResNet-50 architectures on the data set. The findings achieved at this stage are the same as those obtained at the previous stage, which were produced using pre-trained CNN architectures.

Each architecture extracts the same amount of features as the data set's images, which are subsequently concatenated. Finally, these features are optimized and then provided as input to the SVM for prediction of either positive or negative bullying. The attributes of the images in the data set containing bullying and non-bullying photos were extracted using the CNN architectures MobileNetV2, EfficientNetB0, and ResNet50.

These characteristics are derived from the architecture layer preceding the Softmax layer. Data properties are retrieved from the Logits layer in the MobileNetV2 architecture, and from the EfficientNet-B0–model–head–dense–MatMul layer in the EfficientNetB0 architecture. Obtaining these qualities via traditional approaches is a difficult task. Because these traits are manually retrieved, this procedure necessitates specialist knowledge, which is a time-consuming phase (Eroğlu *et al.*, 2021). Deep learning architectures execute feature extraction automatically. The SVM classifier is given features collected automatically in the MobileNetV2, EfficientNetB0, and ResNet50 architectures. The collected results are the second stage of the research. Figure 2 shows how this level works.

The features collected individually in the MobileNetV2, EfficientNetB0, and ResNet50 designs were integrated in the third step of our study. Each architecture's attribute map has 1500×1000 pixels. The size of the feature map obtained after the merging procedure is 1500×3000 pixels. During the merging process, the number of features does not change. The goal is to put the properties of each datum in distinct designs next to one another, rather than under one another. The features of the three architectures are combined in this fashion (Eroğlu *et al.*, 2021). With the suggested model, the concatenated features are exposed to a classification process. The working logic of the suggested model is depicted in Fig. 2.

3.2. SVM classification. The suggested approach is based on the SVM, which has excellent generalization properties and a very accurate paradigm. The main rationale for utilizing the SVM is its ability to solve the overfitting problem that happens in neural networks (Elmezain, 2016; Elmezain and Ibrahem, 2021). Furthermore, it has a risk-minimization structural principle. The SVM has a structure that allows it to run dichotomic classes, especially in higher-dimensional spaces, as well as hypothesize a maximal separation hyperplane. Our approach depends on obtaining the separating hyperplane via maximizing the distance between two parallel hyperplanes as in Fig. 3.

As a result, a good separation based on the largest distance is carried out, which results in a low generalization error of the classifier over the hyperplane.

Assume that the learning data set is represented as $D = \{(x_i, y_i)|x_i \in \mathbb{R}^d, y_i \in \{-1, +1\}\}$, where x is to the observation features, y refers to the labels of the SVM. In addition, the idea deal with the learning problem by providing some instances that make it possible to go beyond the margin constraints (Redi and Merialdo, 2012). The slack parameters are used to articulate prospective margin violations and penalty criteria to prevent them. To train the SVM linearly, the following function is used:

$$f(x) = \mathrm{sign}(w \cdot x + b), \tag{1}$$

where w denotes a weight vector, and x denotes an input sample. The threshold value is b. Using an SVM trainer, the maximum margin for data hyperplanes is considerably increased. The hyper plane margin is maximized to get the shortest distance between the hyperplane and the support vector. Therefore, the SVM is written as follows:

$$\gamma = \frac{2}{\|w\|}. \tag{2}$$

In this case, γ is at the hyperplane's margin. The maximized margin for the hyperplane is shown in Fig. 3, and the input data are split and mapped linearly using the SVM in a high-dimensional domain as in Fig. 4.

The conclusion of the SVM, which gives a high score to all labels, is based on relevance scores computed with respect to two classes with the same values. As a result of the dot product and the kernel trick, the mapping has no effect on the learning time. The SVM as a classifier is the best and strongest against the curse of dimensionality, especially when extracting more features. Furthermore, multiple studies using SVMs have produced acceptable results in a variety of domains. The SVM has the advantage of allowing regression optimization to be corrected through testing and learning processes. Using duality features, the margin and the kernel type, the SVM technique can be restructured in terms of domain difficulties. SVMs are used to overcome problems such as local minima and nonlinearities. Furthermore, SVMs can distinguish between labels and properly segregate them.

3.3. CNN-SVM model. Learning an observation model is an important part of training an SVM. A CNN for modeling observation features is preferred to SVMs

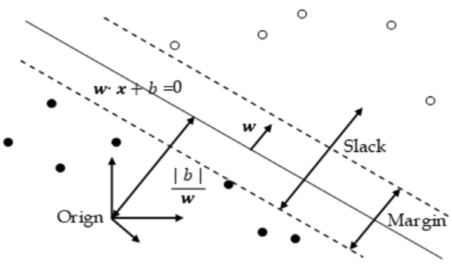

Fig. 3. Hyperplane margin.

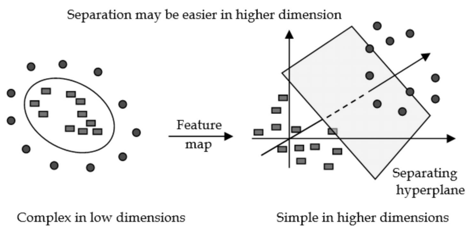

Fig. 4. Kernel function mapping from input data to a richer feature space.

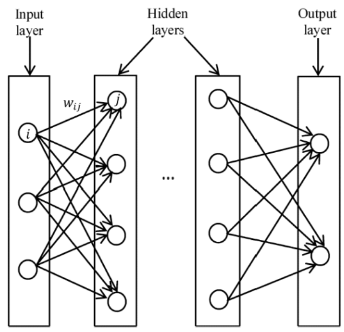

Fig. 5. Graphical representation of a CNN.

in this work. A CNN is a neural network model with an input layer, two or more hidden layers, and an output layer. Figure 5 shows a graphical representation of the CNN model.

The variables are represented by the nodes, and the weighted parameters are represented by the linkages between the nodes. The direction of information flow via the network is indicated by arrows. A CNN contains many more parameters than a standard three-layer ANN since it has a lot of hidden layers in which each hidden layer has more units. Because of the large number of parameters, CNN can automatically extract features for categorization with respect to raw sensor data. The data vector from a frame is supplied into the input units of a CNN. The activation probability y_j of each hidden unit j is computed via the inputs of the previous layer,

$$y_j = \frac{1}{1 + e^{-m_j}}, \quad m_j = b_j + \sum_i y_i w_{ij}, \tag{3}$$

where w_{ij} denotes the weight of the link between units j and i. Unit j has bias b_j. The previous layer's unit y_i has i as its input. The activation probabilities of the units in this layer are then passed on to the next layer as inputs. The output unit j, employs "softmax" to convert the inputs from the previous layer into a classification probability p_j.

$$p_j = \frac{e^{m_j}}{\sum_k e^{m_k}}, \quad m_j = b_j + \sum_i y_i w_{ij}, \tag{4}$$

where k is an all-classes index. The weights and biases of the CNN can be set using either a random initialization strategy or Hinton's pre-training method (Hinton *et al.*, 2006), which entails treating every two surrounding layers as a restricted Boltzmann machine (RBM). After initializing the parameters, the complete network is trained with the information from the labels using the backpropagation algorithm to optimize the parameters.

Following CNN training, the CNN generates the class probabilities $p(Z_t|x_t)$ for an observation x_t at time step t. When the number of classes is N, Eqn. (4) yields $(P_1, P_2, \ldots, P_N)$ for $p(Z_t|x_t)$. The emission probability is estimated by dividing the CNN outputs by the total number of the corresponding classes.

4. Experimental evaluation

The data set used in this study, as well as the model outcomes, are thoroughly analyzed. Moreover, experimental evaluations for the proposed model are explored step by step with figures.

4.1. Data collection. Cyberbullying is defined as any form of harassing, threatening, or degrading language over the Internet.

Here are some of the most intriguing cyberbullying statistics:

(i) The most common kind of online bullying is by mean remarks, which account for 22.5% of all incidents.

(ii) 35% of those polled shared a screenshot of someone's status or image in order to make fun of them.

(iii) When asked why they were bullied, 61% of teenagers said it was because of their physical appearance.

(iv) 56% of those who had been harassed on-line said that they had been harassed on Facebook.

(v) Cyberbullying affects 7 out of 10 young individuals before they reach the age of 18.

About 1200 images connected with cyberbullying were retrieved and analyzed for this study. The data set is collected by the authors and many students. This data set, called the bullying data set, was taken from free Internet websites such as Unsplash and Pixabay, which feature all of the images, designs and HD videos under a Creative Commons Attribution (CC0) license, making it safe for personal and commercial use without having to ask for permission. According to `https://unsplash.com/s/photos/cyberbullying` and `https://pixabay.com/images/search/cyberbullying/`, the collected data set consists of two main classes. These classes are bullying class "Yes (1)" and non-bullying class "No (0)" of cyberbullying. Table 1 shows the label (Yes/No) distribution for the data set. By using the images in these data set, it is aimed to predict the cyberbullying more cheaply and accurately by using the proposed SVM models. Figure 6 shows sample images of each of the classes utilized in this data set for No and Yes classes. The data set used includes 617, and 602 images in Yes and No classes, respectively.

Table 1. Label distribution for the data set.

Labels	Training set	Testing set
Yes	502	115
No	489	113

4.2. Evaluation: Performance metrics. A range of criteria, such as accuracy, precision, recall, and $F1$ score, can be used to assess the usefulness and efficiency of different categorization methods. The outputs of each binary classification model can be classified into four groups: true positive (TP), true negative (TN), false positive (FP), and false negative (FN) (Gad and Hosahalli, 2020; Atlam *et al.*, 2000).

The amount of accurately categorized (TN and TP) predictions made across all types of forecasts is defined as the model accuracy,

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{5}$$

The fraction of items correctly identified as positive out of the total number of items correctly identified as positive is known as precision.

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{6}$$

The number of items accurately classified as positive out of the total number of true positives is referred to as recall.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{7}$$

Precision and recall are used to produce the $F1$ score, with 1 being the best and 0 being the poorest. The $F1$

(a)

(b)

Fig. 6. Samples from our own data set, where the top row (a) includes non-bullying class images "No", while the bottom row (b) contains images of bullying class "Yes".

score is obtained as the harmonic mean of precision and recall,

$$F1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}. \tag{8}$$

4.3. Analysis. The study is mainly focused with images that depict cyberbullying in action. The results of the research were achieved through the use of the Google Colab environment. The computer that was utilized in the study was equipped with a Linux operating system, 16 GB of RAM, and a GUP video card. Performance of the models under consideration was assessed using confusion matrices. Accuracy, sensitivity, specificity, and $F1$-score were all computed individually using confusion matrices.

The outcomes of the first stage of implementation were obtained using seven CNN architectures that had been previously trained. The data set was partitioned into two training and testing subsets, each containing 80% and 20% of data, respectively. The obtained confusion matrices for the employed architectures are provided in Figs. 7 and 8. The MobileNetV2 model classified 171 out of 228 test images correctly, and 57 of them are classified incorrectly as shown in Fig. 7. The average accuracy value of the model was 76.32%. The MobileNetV2 model achieved a precision of 72 % in the Yes class, and 79% in the No class, for a recall of 81% in the Yes class, and 69% in the No class.

On the other hand, the EfficientNetB0 model correctly predicted 213 of the 228 images used for testing, while 15 incorrectly predicted as shown in Fig. 7. The average accuracy value of this model is 93%. The efficientNetB0 model achieved a precision of 96% in the Yes class, and 91% in the No class. The EfficientNetB0 model achieved a recall of 90% in the Yes class, and 97% in the No class. Furthermore, 130 out of the 228 photos included in the InceptionResNetV2 model's testing set were correctly classified, whereas the remaining 98 were wrongly classified. In the model, the average accuracy value attained is 57%. In the Yes class, the model produced a 57% $F1$-score, whereas in the No class, it achieved a 57 % $F1$-score.

As demonstrated in Fig. 8, the NASNetLarge model accurately predicted 187 out of 228 photos used in the test, whereas 41 photos were wrongly predicted. The model's average accuracy is 82% . The NASNetLarge model had a precision of 79% in the Yes class and 861% in the No class. In addition, in the Yes class, the NASNetLarge model yielded a $F1$-score of 83%, whereas in the No class, it produced a $F1$-score of 81%. Finally, in the ResNet50 model, 206 out of 228 photos used for testing were correctly classified, whereas 22 were wrongly classified. In the model, the average accuracy value attained is 90%. In the Yes class, the model produced a 90% $F1$-score, whereas in the No class, it attained a 91% $F1$-score.

The second phase of our study was then completed. The top four models MobileNetV2, Xception, EfficientNetB0, and ResNet50 architectures were used to extract features. The number of features extracted and concatenated by each architecture is the same as the number of photos in the data set. The SVM classifier receives the characteristics obtained in each architecture and concatenates them. The SVM classifier was used to identify these features after they had been optimized. 219 out of 228 photos were successfully predicted by the suggested model, while 9 were incorrectly predicted. In

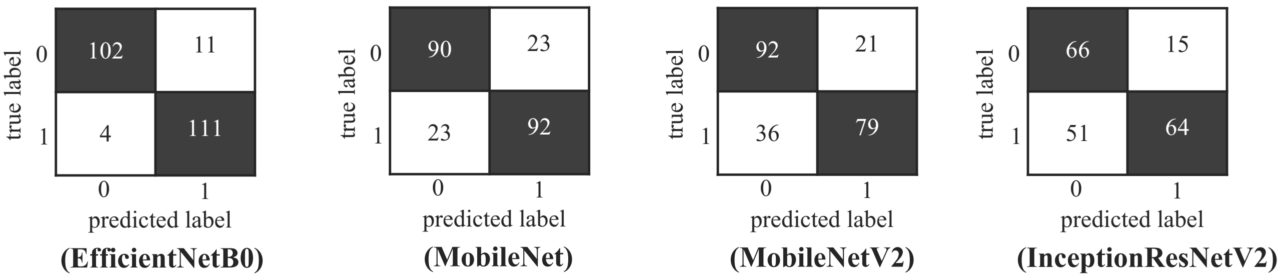

Fig. 7. Confusion matrix for EfficientNetB0, MobileNet, MobileNetV2 and InceptionResNetV2 models.

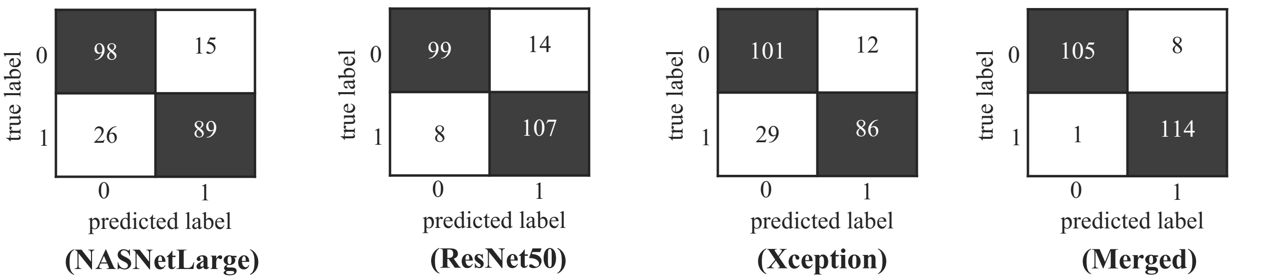

Fig. 8. Confusion matrix for NASNetLarge, ResNet50, Xception and Merged + Softmax.

the "Yes" class, the model had a precision of 99% , and in the "No" class, it had a precision of 93%. The accuracy of the combined models + the SVM classifier was 96%. The suggested performance criteria are listed in Table 2. The merged model had the greatest accuracy value of all the models considered.

Visualizing the performance of a machine learning model is a straightforward method of making sense of the data that comes out of the model. Using this information, we can determine which modifications are necessary to the model parameters or hyperparameters that impact the machine learning model. The number of nodes per layer and the total number of layers in the neural network are hyperparameters that can have a significant impact on the model performance. Visualizing the fitness of the training and validation sets can assist in optimizing these parameters and in developing a more accurate model. One of the most challenging aspects of any deep learning technique is generalizing the model so that it is good at predicting reasonable results with additional data. Visualizing the relationship between training and validation accuracies across a number of epochs is a useful tool for determining whether or not the model has been trained properly. The accuracy curve is also one of the most commonly used curves to understand the behavior of neural networks. A curve with both training and validation accuracies is more meaningful. Figure 9 shows the relationship between the training and validation accuracies as a function of the number of epochs.

5. Discussion

Bullying has always taken the form of stronger individuals preying on the weaker ones, and with the advancement of technology, it has now migrated to the Internet as well. Bullying is a social problem that has captured the attention of educators and policymakers in a number of states in recent years, particularly in India, Brazil, and United States. Additionally, bullying is a situation defined by the repeated use of aggressive verbal or physical abuse towards one or more peers by one or more individuals. Bullying has a tremendous influence on the lives of many persons. According to cyberbullying data, the most widely utilized platform for cyberbullying is Instagram, followed by Facebook and Snapchat. Cyberbullying is a typical occurrence for people who play online multiplayer games. Although YouTube has a large number of users, approximately a tenth of those individuals have reported being a victim of cyberbullying on the website thus far.

Cyberbullying victims have low self-esteem, sadness, and heightened social anxiety, with many considering suicide as a result. Consequently, cyberbullying has been associated with a number of undesirable effects such as drug abuse, eating disorders and alcoholism, as well as lower levels of academic performance. All of this serves to raise awareness about the hazards of cyberbullying and the importance of taking action to combat it. These serious cyberbullying statistics demonstrate how big the destructive impact of bullying is, and how critical it is to take a decisive action to put an end to it once and for all. It is essential for the general

Table 2. Classification report of the proposed merged model.

	Precision	Recall	F1-score	Support
Yes	0.99	0.93	0.96	113
No	0.93	0.99	0.96	115
Accuracy			0.96	228
macro avg	0.96	0.96	0.96	228
weighted avg	0.96	0.96	0.96	228

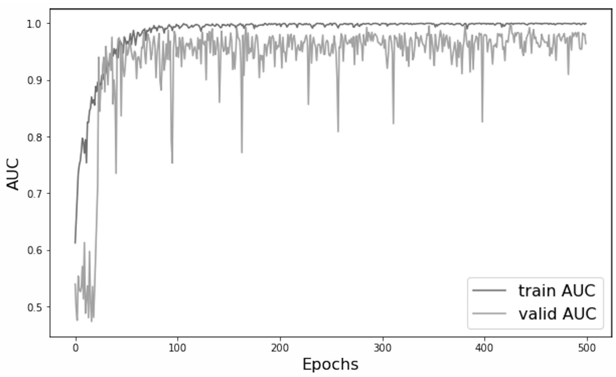

Fig. 9. Accuracy curve of the proposed model.

health of those who are being targeted. People who were cyberbullied experienced social anxiety in 41% of cases, anxiety in 37% of cases, and even suicidal thoughts in 26% of cases. The results of a 2017 study conducted by Statistic indicated that a significant percentage female victims of cyberbullying suffer from a variety of negative consequences that affect their mental health and general well-being. Cyberbullied women reported feeling helpless in their abilities to respond to the abuse, and a whopping 66% reported being unable to sleep properly as a result. Another notable negative consequence of cyberbullying is a decrease in self-confidence, which was observed in 61% of young women respondents to be the case.

Professionals make every effort to address or prevent this occurrence. Different professionals, including family members or society and police officers, are collaborating everyday to help bullied persons overcome their issues. Numerous social organizations have been established to provide assistance and actual solutions for a large number of bullied individuals. However, the organizations make big efforts to simplify and reduce the cost of these solutions, as well as to enhance the validity of the results. While the majority of people are aware of the facts about bullying and cyberbullying, only a minority are aware of how to handle it. Those who observe it seldom intervene, mostly because they are afraid of the consequences of being involved themselves. It is common for parents to be unaware that their kid is being cyberbullied because most youngsters feel it to be a typical occurrence and do not want to inform their parents about it. On social media sites, most students seek to avoid cyberbullying by blocking the bullies, and so far, this appears to be the most effective method.

Websites like `StopBullying.gov` have recently gained popularity, offering victims of bullying the tools they need to seek help and reclaim their lives. Many children are aware of the dangers of cyberbullying. One way to avoid cyberbullying is to share as little personal information online as possible. Bullies are less likely to harass someone they do not know to exist. State and policy implementations impact cyberbullying rules in the US. Cyberbullying and online harassment are crimes in 48 states, while criminal consequences are in place in 44 states for those who participate in it. Young people believe social media platforms should do more to address cyberbullying. While blocking bullies on social media may seem like an ideal option, blocking them does not stop them from spreading rumors and harassing others online.

Using artificial intelligence techniques, it will be able to conduct the classification of the real images collected from different locations. This work utilized deep learning architectures to build a bullying classification model. The proposed technique achieved a 96% accuracy, which was impressive. Using pre-trained MobileNetV2, ResNet50, and EfficientNetB0 models, the suggested technique achieved the needed. The features taken from the input image are then classified using the SVM classifier. The trained ResNet50, EfficientNetB0, MobileNet, Xception, NASNetLarge, MobileNetV2 and InceptionResNetV2 architectures were merged to produce the features. When these combined features were classified in the SVM classifier, the best results were obtained. Table 3 shows the accuracy results obtained in the first and second phases of the experiments. In terms of accuracy, the results produced by the combined model outperform those obtained by pre-trained models.

Complications exist on two levels: during training and during testing. We estimate the vector w and the bias b by solving a quadratic problem during training for linear SVMs, and prediction is linear in the number of features and constant in the amount of training data during testing. The number of support vectors (which can be constrained by the training set size multiplied by the training set error rate) and the number of features are determined during training, and the hardness of testing is determined by the number of support vectors (which can be constrained by the training set size multiplied by the training set error rate) and the number of features.

The training data set in our proposed approach is 991 images with a 0.009 error rate, hence the time complexity is O(991$\times$ 0.001).

The values in Table 3 are ordered from highest to lowest accuracy. The suggested merging technique gets the best results in terms of accuracy, recall, precision, and $F1$-score.

Additionally, when the suggested model is compared with other studies in the literature, as shown in Table 6, successful findings are achieved.

Figure 10 illustrates a real-world example of the suggested model, as well as the predicted probabilities for the Yes and No classes. From Fig. 10, it is clear from the output of one of the images that class Yes is predicted with a probability of 0.007% and class No is predicted with a probability of 99.993%. However, the probability of 1 occurs even when there is no indication of cyberbullying in a image, which is acceptable for images that include an image of some desired class. In the bottom row example, using the probability distribution, class Yes is predicted with a probability of 100% and class No is predicted with a probability of 0.0% in Fig. 10.

6. Conclusion

Cyberbullying is a crime that occurs when someone expresses anger, envy, filthy language, or insulting comments in writing, speech, or gestures. While many countries are attempting to enact anti-bullying legislation, an effective preventative solution has yet to be developed, and many citizens are dissatisfied with how bullying is dealt with. While many countries are attempting to enact anti-bullying legislation, an effective preventative solution has yet to be developed, and many citizens are dissatisfied with the way bullying is handled in their countries.

In this study, a hybrid classified model based on transformers in conjunction with an SVM model is proposed to predict whether or not bullying takes place. The accuracy rate of the proposed merged models with the SVM classifier achieved 96.05%. Furthermore, the classification precision of the proposed merged model is 99% in the bullying class, and 93% in the non-bullying class. According to these results, bullying has a negative impact on students' academic performance. Moreover, it helps stakeholders take an appropriate action against anti-bullying and raising community awareness of the problem. Future works will focus on using Twitter texts with Google form questionnaires for classifying cyberbullying and how to stop it.

Acknowledgment

We would like to show our gratitude to and acknowledge Dr. Zohair Malki, the previous dean of the College of Computer Science and Engineering at Taibah University. We want to thank him for sharing his extensive knowledge with us during preparation of this paper and for his comments that greatly improved the manuscript. We dedicate this paper to his hard work.

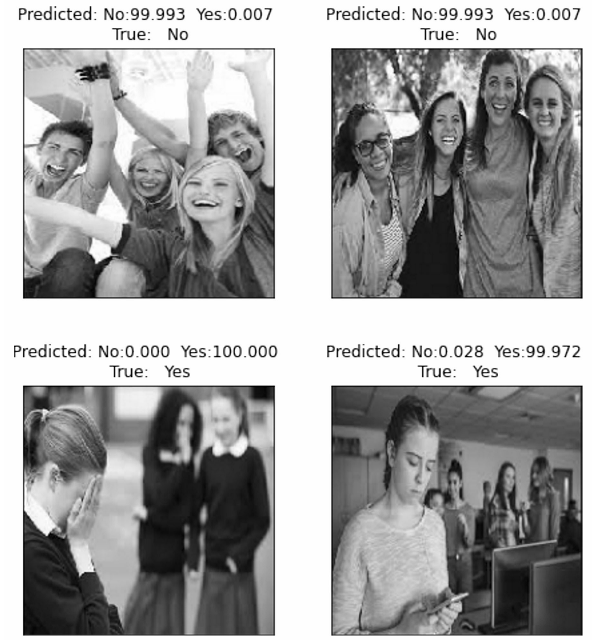

Fig. 10. Real-world example of the proposed model, where the top row includes some non-bullying images, while the bottom row contains images of bullying.

References

Abaido, G.M. (2019). Cyberbullying on social media platforms among university students in the United Arab Emirates, *International Journal of Adolescence and Youth* **25**(1): 407–420.

Alim, S. (2015). Analysis of tweets related to cyberbullying, *International Journal of Cyber Behavior, Psychology and Learning* **5**(4): 31–52.

Atlam, E.-S., Ewis, A., El-Raouf, M.A., Ghoneim, O. and Gad, I. (2022). A new approach in identifying the psychological impact of COVID-19 on university student's academic performance, *Alexandria Engineering Journal* **61**(7): 5223–5233.

Atlam, E.-S., Fuketa, M., Morita, K. and Ichi Aoe, J. (2000). Similarity measurement using term negative weight and its application to word similarity, *Information Processing and Management* **36**(5): 717–736.

Blachnik, M. (2019). Ensembles of instance selection methods: A comparative study, *International Journal of Applied Mathematics and Computer Science* **29**(1): 151–168, DOI: 10.2478/amcs-2019-0012.

Boulton, M.J. and Underwood, K. (1992). Bully/victim problems among middle school children, *British Journal of Educational Psychology* **62**(1): 73–87.

Cleetus, L., Sukumar, A.R. and Hemalatha, N. (2021). Computational prediction of disease detection and insect identification using xception model, *bioRxiv*, DOI: 10.1101/2021.08.10.455608, (preprint).

Cortis, K. and Handschuh, S. (2015). Analysis of cyberbullying tweets in trending world events, *Proceedings of the 15th*

Table 3. Results obtained in the study.

	Accuracy	Precision	Recall	F1
Merged	0.960526	0.930687	0.991304	0.960478
ResNet50	0.929825	0.900797	0.930435	0.929825
EfficientNetB0	0.925439	0.878858	0.973913	0.925439
Xception	0.815789	0.770019	0.773913	0.815789
MobileNet	0.802632	0.753696	0.765217	0.802632
NASNetLarge	0.780702	0.716778	0.817391	0.780702
MobileNetV2	0.763158	0.710868	0.721739	0.763158
InceptionResNetV2	0.627193	0.589586	0.513043	0.627193

International Conference on Knowledge Technologies and Data-driven Business, Graz, Austria, Article no. 7, pp.1–8

Delprato, M., Akyeampong, K. and Dunne, M. (2017). The impact of bullying on students' learning in Latin America: A matching approach for 15 countries, *International Journal of Educational Development* **52**: 37–57.

Devidas, S., Rao, Y.V.,S. and Rekha, N.R. (2021). A decentralized group signature scheme for privacy protection in a blockchain, *International Journal of Applied Mathematics and Computer Science* **31**(2): 353–364, DOI: 10.34768/amcs-2021-0024.

Elmezain, M. (2016). 3D fingertip detection on multi-class support vector machine based hand gesture recognition, *Kasmera* **44**(8): 83–89.

Elmezain, M. and Ibrahem, H.M. (2021). Retrieving semantic image using shape descriptors and latent-dynamic conditional random fields, *The Computer Journal* **64**(12): 1876–1885.

Elmezain, M., Othman, E.A. and Ibrahim, H.M. (2021). Temporal degree-degree and closeness-closeness: A new centrality metrics for social network analysis, *Mathematics* **9**(22): 2850.

Emine Cengil, A.C. (2019). Multiple classification of flower images using transfer learning, *International Artificial Intelligence and Data Processing Symposium (IDAP), Malatya, Turkey*, pp. 1–6.

Eroğlu, Y., Yildirim, M. and Çinar, A. (2021). Convolutional neural networks based classification of breast ultrasonography images by hybrid method with respect to benign, malignant, and normal using mRMR, *Computers in Biology and Medicine* **133**, Paper ID: 104407.

Fekkes, M. (2006). Do bullied children get ill, or do ill children get bullied? A prospective cohort study on the relationship between bullying and health-related symptoms, *Pediatrics* **117**(5): 1568–1574.

Gad, I. and Hosahalli, D. (2020). A comparative study of prediction and classification models on NCDC weather data, *International Journal of Computers and Applications* **44**(5): 414–425.

Harjoseputro, Y., Yuda, I.P. and Danukusumo, K.P. (2020). MobileNets: Efficient convolutional neural network for identification of protected birds, *International Journal on Advanced Science, Engineering and Information Technology* **10**(6): 2290.

Hark, K. and Turgut Özal, E.S. (2019). Extractive text summarization via graph entropy (Çizge Entropi ile Çıkarıcı Metin Özetleme), *2019 International Artificial Intelligence and Data Processing Symposium (IDAP), Malatya, Turkey*, pp. 1–5.

Hellsten, L., Crespi, I., Hendry, B. and Fermani, A. (2021). Extending the current theorization on cyberbullying: Importance of including socio-psychological perspectives, *Italian Journal of Sociology of Education* **13**(3): 85–110.

Hinton, G.E., Osindero, S. and Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets, *Neural Computation* **18**(7): 1527–1554.

Ibrahem, H.M., Elmezain, M. and Shoman, S. (2020). Adaptive image enhancement approach based on double-plateaus histogram, *Journal of Theoretical and Applied Information Technology* **98**(10): 1675–1685.

Kibriya, S., Xu, Z.P. and Zhang, Y. (2015). The impact of bullying on educational performance in Ghana: A bias-reducing matching approach, *2015 AAEA & WAEA Joint Annual Meeting, San Francisco, USA,* Paper no. 205409.

Kosciw, J.G., Palmer, N.A., Kull, R.M. and Greytak, E.A. (2012). The effect of negative school climate on academic outcomes for LGBT youth and the role of in-school supports, *Journal of School Violence* **12**(1): 45–63.

Le, A.T., Miller, P.W., Heath, A.C. and Martin, N. (2005). Early childhood behaviours, schooling and labour market outcomes: Estimates from a sample of twins, *Economics of Education Review* **24**(1): 1–17.

Longobardi, C., Thornberg, R. and Morese, R. (2022). Cyberbullying and mental health: An interdisciplinary perspective, *Frontiers in Psychology* **12**, Article ID: 827106.

Malki, Z., Atlam, E.-S., Hassanien, A.E., Dagnew, G., Elhosseini, M.A. and Gad, I. (2020). Association between weather data and COVID-19 pandemic predicting mortality rate: Machine learning approaches, *Chaos, Solitons & Fractals* **138**, Article ID: 110137.

McGuire, L. and Norman, J. (2018). Parents coping with cyberbullying: A bioecological analysis, *in* M. Campbell and S. Bauman (Eds), *Reducing Cyberbullying in Schools*, Elsevier, Amsterdam, pp. 61–72.

Mingxing Tan, Q.V.L. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks, *International Conference on Machine Learning, Long Beach, USA*, pp. 6105–6114.

Ponzo, M. (2013). Does bullying reduce educational achievement? An evaluation using matching estimators, *Journal of Policy Modeling* **35**(6): 1057–1078.

Radhika, K., Devika, K., Aswathi, T., Sreevidya, P., Sowmya, V. and Soman, K.P. (2020). Performance analysis of NASNet on unconstrained ear recognition, *in* M. Rout *et al.* (Eds), *Nature Inspired Computing for Data Science*, Springer International Publishing, Cham, pp. 57–82.

Redi, M. and Merialdo, B. (2012). A multimedia retrieval framework based on automatic graded relevance judgments, *International Conference on Multimedia Modeling, Klagenfurt, Austria*, pp. 300–311.

Rui Zhao, Y.Z. and Chen, Z. (2019). Deep learning and its applications to machine health monitoring, *Mechanical Systems and Signal Processing* **115**(3): 1568–1574.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. and Chen, L.-C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks, *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA,* pp. 4510–4520.

Sarzosa M. and Urzúa S. (2015). Bullying among Adolescents: The Role of Cognitive and Non-Cognitive Skills, *National Bureau of Economic Research* **5**(4): 10–14.

Setiana, D. and Besar, N. (2021). ICT emerging technology impact within learning ecosystem cyberbullying among students, *in* P. Ordóñez de Pablos *et al.* (Eds), *Handbook of Research on Analyzing IT Opportunities for Inclusive Digital Learning*, IGI Global, Hershey, pp. 154–171.

Tripathi, K., Gupta, A.K. and Vyas, R.G. (2020). Deep residual learning for image classification using cross validation, *International Journal of Innovative Technology and Exploring Engineering* **9**(4): 1525–1530.

Zhang, H., Liu, C., Zhang, Z., Xing, Y., Liu, X., Dong, R., He, Y., Xia, L. and Liu, F. (2021). Recurrence plot-based approach for cardiac arrhythmia classification using inception-ResNet-v2, *Frontiers in Physiology* **12**: 1–12, Paper ID: 648950.

Zohair, Elsayed, Ewis, A., Dagnew, G., Reda, A., Elmarhomy, G., Elhosseini, M.A., Hassanien, A.E. and Gad, I. (2021). ARIMA models for predicting the end of COVID-19 pandemic and the risk of a second rebound, *Neural Computing and Applications* **33**(7): 2929–2948.

Mahmoud Elmezain was born in Egypt. Between 1997 and 2004 he worked as a demonstrator in the Department of Statistics and Computer Science, Tanta University, Egypt. He received his MS degree in computer science from Helwan University, Egypt, in 2004. He also holds a PhD degree in computer science from the Institute for Electronics, Signal Processing and Communications at the Otto von Guericke University of Magdeburg, Germany. His work focuses on image processing, pattern recognition, human-computer interaction, action recognition, machine learning and deep learning. Doctor Elmezain is the author of more than 100 articles in peer-reviewed international journals and conferences.

Amer Malki received his PhD in information sciences, specialising in telecommunication and networking, from the School of Computing and Information (SCI), University of Pittsburgh, USA, in 2019. He obtained his MS degrees in telecommunication system management and information assurance from Northeastern University in Boston, USA, in 2009. He received his BSc degree in electrical engineering from King Fahd University of Petroleum and Minerals in Dharan, Saudi Arabia, in 1998. He is now an assistant professor at the College of Computer Sciences and Engineering of Taibah University, Yanbu branch. Doctor Amer has worked in different sections in the telecommunications department for the government of Saudi Arabia. His areas of interest include, network science, telecommunications policies, cyber security (information assurance), cryptography, blockchain, natural language processing, machine learning, and pattern recognition.

Ibrahim Gad graduated from Tanta University in 2006 with a BS in computer science. In 2006–2014, he worked as an instructor at the Department of Computer Science, Faculty of Science, Tanta University. He received his MS degree in computer science in 2014 from the Department of Computer Science, Faculty of Science, Ain Shams University, Cairo, Egypt. From 2014 to 2020, was a teaching assistant in the Department of Computer Science, Faculty of Science, Tanta University. From 2016 to 2020, he was a research scholar in the Department of Computer Science, Faculty of Science, Mangalore University. Since 2020, he has been an assistant professor in the Department of Computer Science, Faculty of Science, Tanta University. His research interests are in data mining, big data, machine learning, deep learning, and natural language processing.

El-Sayed Atlam received his BS and MS degrees in mathematics from the Faculty of Science, Tanta University, Egypt, in 1990 and 1994, respectively, and his PhD degree in information science and intelligent systems from the University of Tokushima, Japan, in 2002. From 2003 to 2005 he held a postdoctoral fellowship in the Department of Information Science and Intelligent Systems, Tokushima University. Doctor Atlam is a board member of the *Computer Algorithm Series* of the IEEE Computer Society Press (CAS) and the Egyptian Mathematical Association (EMA). His research interests include information retrieval, natural language processing, document processing and sentence retrieval from huge text data bases and morphological analysis.

Received: 6 December 2021
Revised: 1 February 2022
Re-revised: 3 March 2022
Accepted: 7 March 2022