ISSN: 2584-0495 Vol. 1, Issue 3, pp. 173- 183

International Journal of Microsystems and IoT

ISSN: (Online) Journal homepage: https://www.ijmit.org

Hybrid shark and bear smell optimization algorithm with deep neural network for sentiment analysis.

Nainika Kaushik, Manjot Kaur Bhatia

Cite as: Nainika Kaushik, & Manjot Kaur Bhatia. (2023). Hybrid shark and bear smell optimization algorithm with deep neural network for sentiment analysis. International Journal of Microsystems and IoT, 1(3), 173-183. https://doi.org/10.5281/zenodo.8365535

Published online: 21 August 2023.

DOI: https://doi.org/10.5281/zenodo.8365535

International Journal of Microsystems and IoT
Vol. 1, Issue 3, pp. 173- 183; DOI: https://doi.org/10.5281/zenodo.8365535

Hybrid shark and bear smell optimization algorithm with deep neural network for sentiment analysis.

Nainika Kaushik, Manjot Kaur Bhatia

Jagan Institute of Management Studies, Rohini Sector 5, Delhi, India (Corresponding author e-mail: nainika.kaushik@jimsindia.org)

ABSTRACT

Nowadays, Twitter is a well-liked media platform system where people can publish any browser material. Since the trends found will be so helpful to telemedicine in a specific way, this openly accessible user information is indeed essential. These are some of the uses is the computerized detection of psychological issues, such as anxiety. Earlier research on accurately detecting depressed individuals on social sites has mostly depended upon consumer behaviors, language skills, and interpersonal relationships. The drawback is that those same systems are developed on even a variety of unrelated data that may not be essential for identifying a depressed individual. Additionally, the performance and efficiency of the system as a whole are negatively impacted by these contents. Designers suggest a unique multimodal concept, it's a combination of recurrent neural network and attention-based residual network categorization unit with decision-level fusion algorithms that address the limitations of the current artificial depressive detection techniques. The Hybrid Shark and Bear Smell Optimization Algorithm extract features from audio, video, and text.

Keywords:
Bear smell optimization algorithm; Classification; Deep learning; Deep neural network; Depression; Feature extraction; Fusion; Sentiment Analysis; Shark optimization algorithm; social media; Twitter.

1. INTRODUCTION

Currently, mental stress is endangering population livelihoods. Ever more individuals experience sadness because of the quickening busy lifestyle. Almost fifth of the people, according to a global poll cited by new business, have seen a substantial increase in stress over the past two years [1]. Although anxiety is a non-clinical yet normal occurrence in daily life, extreme and persistent depression could be detrimental to a person's physical and mental well-being. According to published research studies, chronic depression has been linked to a variety of illnesses, including chronic depression, sleeplessness, and other conditions. [2] Additionally, the Centers for Disease Control and Prevention estimates that distress is a big contributor to suicide [3], that's become the leading cause of morbidity and mortality people.

Each user shares a diverse range of online comments, simply including those that are treated with medication; it is crucial to identify sadness early on before everything develops into serious issues. The major methods used in current psychological emotion recognition are in-person interviews, self-report questionnaires, or wearable sensors. Traditional approaches, on the other hand, are reflexive but typically time-consuming, labor-intensive, as well as hysterical [4]. Do any quick, effective ways to identify anxiety? Both lifestyles as well as academics in medicine are transforming as a consequence of the rising of social media. Ever more individuals were inclined to discuss its everyday happenings as well as feelings that socialize with people over social media thanks to the proliferation of sites like Twitter [18] and Chinese Weibo [5]. Such social media marketing offers the chance for depicting, assessing, modelling, as well as resource extraction [19] user activity textured surfaces along through massive social connections, and certain relevant information could indeed discover the conceptual underpinning in behavioral science because they closely resemble subscribers' actual regions and feelings together in timely basis. Previous research [6,7] has shown that this is possible to use social media to build medicine, specifically in the context of depressive identification. Posting on twitter Material Predicated Depressive Identification Has Disadvantages. Firstly, individuals may not always effectively articulate their depressive emotions in posts because messages via platforms such as twitter or Facebook are constrained toward a capacity of 150 words. According to a new Pew Research Centre, person with greater mental anxiety would be less engaged on social media. Owing to the inherent small datasets and vagueness issues caused by these occurrences, microblogging material depression diagnosis effectiveness could suffer. Online Social Transactions Provide Helpful Cues for Stress Detection. Two fascinating findings range from social behavioral sciences. The first would be attitude diffusion [8]: throughout social relationships, a negative mood could spread by one individual to the next. The next is verbal reverberation [9]: it is normal for people to imitate their mannerisms and emotions of others. Such findings encourage everyone to widen the definition of our message on twitter analysis but include subsequent social interactions, such as discussions and reposting, inside the identification of consumer sadness. In fact, this might help to lessen the semantic gap issue for an individual user. The authors [13] offer a unique mathematical method for something like the automatic identification of depression who first chooses the most pertinent information from the succession among all user twitter posts using a hybrid extractive and abstractive

summarizing technique, producing more precise and pertinent information. The information is being sent to our cutting-edge learning system, which combines attention-enhanced Gated Recurrent Units (GRU) models using Convolutional Neural Networks (CNN) to produce estimation evidence which are superior to those of established baselines. To identify anxiety amongst teenagers, [14] generally uses processing of Sina Weibo statistics. Firstly, it is significant to gather text data from Online users who are students, then turn that textual data into predictive modelling data input. To extract features, neural network is employed. The input data are classified using the deep integrated support vector machine (DISVM) technique [20], which enables depression to be recognized. The priori knowledge is stabilized and indeed the reliability of diagnosing depression is improved to some degree using above mentioned method. The researcher [15] suggested a crawler-based technique for retrieving useful information from the web. The data is then pre-processed. Particle swarm optimization, support vector machine classification, and k-means clustering are all used in the comparison.

Individuals that are depressed typically behave differently on social media, creating rich metadata that is frequently utilized to identify different components. Not every one of these characteristics, nevertheless, have been connected to depression-related traits. Supervised learning was subsequently used to identify sadness on social networking sites, and the results were massively better than those of typical machine learning techniques [10]. While latest evidence demonstrated the efficacy of classifiers [16] for stress detection, most current classification algorithms will not provide comprehensibility for anxiety forecasting; as a result, with these predictions are opaque to humans, that decreases confidence in the neural network models. An open to interpretation model enhances comprehension and offers clear insight into whether a trained model might be improved [11,12]. Therefore, besides employing multi-modal information from different behaviors of the depressed individual in media platforms, researchers intend to construct an interpretable deep learning-based approach for anxiety identification in terms of getting the proper trustworthiness and explain the reasoning behind judgment. In addition to the underlying characteristics generated by linguistic qualities, researchers observe that the patterns of twitter, or the tweeting timelines, offer a critical indication describing the evolution of depressed user sentiment [17]. Throughout order to improve the identification of sad consumers using multi-modal information, we suggest a recurrent neural network (RNN)_ResNet.

2. PROPOSED METHODOLOGY

The overall block diagram for the proposed sentiment analysis using multimodal dataset is shown in Fig.1.Initially the twitter social media consists of audio, video and text based multi modal dataset. This dataset is hence pre-processed by removing the stop words, URL's extraction, special characters and symbols removal for text modality, whereas Degree of Gradient method is adopted for audio and video modality. Then data annotation and vectorization are done for feature extraction step using Hybrid Shark and Bear Smell Optimization Algorithm (HSBOA). The extracted features are ready for decision level fusion process. Finally, the fusion data is classified using RNN+Att_ResNet.

2.1 Pre-processing of data

Fixed length stemming is what is done in this stage. The first n letters are kept in this primitive chopping while the remaining ones are thrown away. For example, when length n = 1 (FIX1), the first letter alone is maintained. Here, FIX1 is termed ultra-stemming. Moreover, it is considered that this can enhance the quality of summaries generated. Actually, the two versions discussed below are used in the experiments:

FIXn: The first n letters are maintained. Words with length < n are not disturbed. Thus, producing data sparseness problem (over stemming effect).

SFIXn: The first n letters are maintained. Words with length < n are omitted. Hence, it is ensured that noise is omitted to some extent but there even occurs a loss of some valuable information.

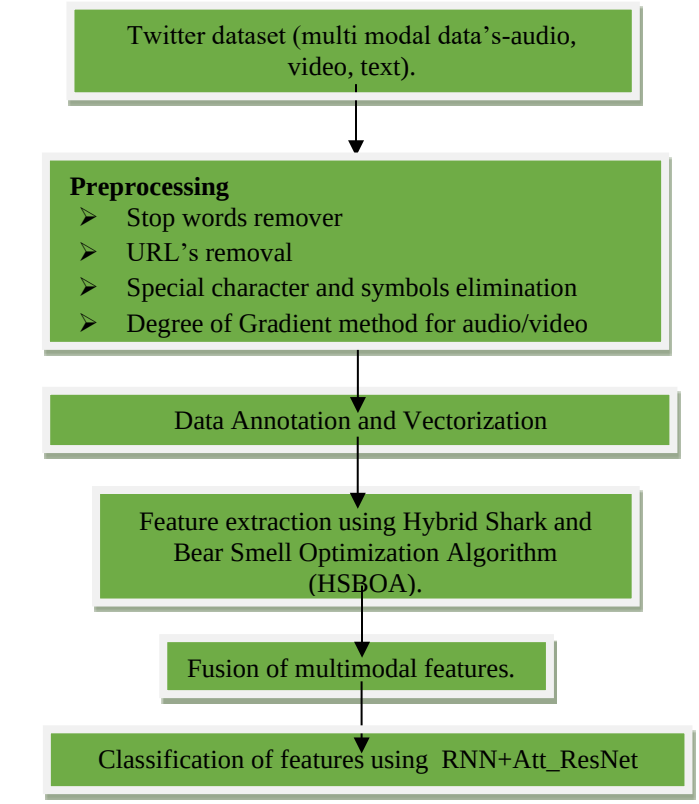

Fig.1 Block diagram for the proposed sentiment analysis using multimodal dataset

Difference Of Gaussian (DOG) Filtering- Difference of Gaussians is a grayscale image enhancement algorithm used in computer vision that includes reducing one blurry

version of an initial grayscale picture from another that is less blurry. The source grayscale image is convolved with Gaussian kernels with varying standard deviations to create the blurred images. When a Gaussian kernel is used to blur a picture, only high-frequency spatial information is suppressed. The spatial information between both the frequency ranges maintained in the two blurred pictures can be kept by removing one image from the other. As a result, the difference between Gaussians is a band-pass filter that eliminates all but a few sub bands from the source grayscale image. The Difference of Gaussian (DOG) technique for contrast enhancement can be used to make edges and other details in a digital picture more visible. The Difference of Gaussians technique eliminates high frequency information, which frequently contains randomness, therefore this method may be found to be very effective for processing photos with a high level of noise. The following is the DOG impulse response:

$$DOG(x,y) = \frac{1}{2\pi\sigma 1^2}.e^{\frac{x^2+y^2}{2\sigma 1^2}} - \frac{1}{2\pi\sigma 2^2}.e^{\frac{x^2+y^2}{2\sigma 2^2}} \quad (1)$$

where 1.0 and 2.0, respectively, are chosen as the default values for the variables 1 and 2. The contrast must be increased in the following phases since this impact causes the total contrast generated by the procedure to decrease. Contrast equalization- The picture intensities are rescaled in the last step of the pre-processing procedure. Since the signal frequently comprises, extreme values caused by highlights, tiny dark areas like noses, rubbish at the picture borders, etc., it is crucial to utilize a powerful estimator.

For this, one might (for instance) use the median of the signal's absolute value, but a straightforward and quick estimate based on the following two stages is preferred:

$$I(x,y) = \frac{I(x,y)}{(mean(min(c,I(x^F,y^F)^\delta)*^1/\delta} \quad (2)$$

$$I(x,y) = \frac{I(x,y)}{(mean((c,I(x^F,y^F)^\delta)*^1/\delta} \quad (3)$$

Here, a threshold is used to terminate big values after the very first phase of normalization, and the mean is across the entire (unmasked part of the) picture is a very compressive exponent that lowers the effect of large values.

$$DS_i^j = \begin{cases} MG_i\ (t - t\_inhale) + DS_i^{t\_inhale} & t\_inhale \le t \le t\_exhale \\ DS_i^{t\_exhale}\ exp\left(\frac{t\_exhale - t}{\varepsilon\ exhale}\right) & t\_exhale \le t \end{cases} \quad (6)$$

where t_ exhale, t_inhale and ε exhale are, respectively, exhalation time, inhalation time, and an exhalation time constant. In the proposed method, the total duration of a breathing cycle is equal to k or the length of the ith odour, and the odour components are separated into two groups according to the exhale and inhalation timings.

2.2.2 Odour absorption:

Now, $MG = \{MG_1, MG_2, \ldots MG_i, \ldots MG_n\}$ has an input (combined their information) for the ith mitral, which contains the receptor sensitivities, odour absorption, and identification. DS_i^j=0 demonstrates that there is no smell in the air before the following inhale. Calculations for the non - negative real set MG include:

2.2 Text features extraction using Hybrid Shark and Bear Smell Optimization Algorithm (HSBOA)

Shark Smell Optimization (SSO) has depended on the shark's aptitude since it excels at employing a powerful smell sense in a short amount of time to grab prey. The primary function of the bear's olfactory bulb, which is far bigger than that of other mammals, is to transmit scent data from the nostril to the brain. This technique, called Bear Scent Optimization (BSO), uses the bear's sense of smell to help it discover food thousands of kilometers away (or global solution in optimization). The numerical method based on the ability to smell provides a potent method to discover the target because bears cannot see food at a great distance. By combining these two algorithms we can obtain better fitness value for feature extraction process.

2.2.1 Initialization process:

Within the search area, a population of the SSO algorithm's first solution must be produced at random. Each of these answers corresponds to an odour particle that indicates a potential Shark location at the start of the search procedure. In (4) and (5), respectively, the initial solution vector is displayed, with x_i^1=ith starting location of the population vector and NP=population size.

$$X^1 = [x_1^1, x_2^1, \ldots . x_{NP}^1] \quad (4)$$

The corresponding optimization issue may be written as follows:

$$x_i^1 = [x_{i,1}^1, x_{i,2}^1, \ldots . x_{i,ND}^1] \quad (5)$$

where ND referred for the decision variable number and $x_{(i,j)}^1$represented the jth dimension of the shark's ith location. Because everything in the surroundings has a distinctive fragrance, the bear used the Bear Smell Optimization approach to make his nose absorb several scents so that each one displayed a position for movement. A lot of them are designated as local solutions, as you will see. The ultimate and most comprehensive answer is the distinct aroma of the desired dish. Let DS_i^jrepresent the jth odour component in the ith odour in accordance with the glomerular carry out activities and respiratory function in a sniff cycle. Using mathematical formulas, one obtains:

$$MG_i\ (O_i) = \frac{1}{k}\sum_{j=1}^{k} f(oc_i^j), f(oc_i^j) = \begin{cases} 1, T_1 \le oc_i^j \\ 0\ T_1 > oc_i^j \end{cases} \quad (7)$$

where T1 is a threshold variable and sets based on average value of odours information and k signifies the odour duration in ith odour. The dynamics of the brain's granular and mitral layers.

$$X = -H_0\omega_y(Y) - a_x X + \sum L_0\omega_y\ (X) + DS \quad (8)$$

$$Y = W_0\omega_y\ (x) - a_y\ Y + DS \quad (9)$$

Where $X = \{x_1, x_2, \ldots x_n\}$. $Y = \{y_1, y_2, \ldots y_n\}$ are, respectively, the mitral and granule cells activity. $DS = \{ds_1, ds_2, \ldots ds_n\}$ and $DS_c = \{ds_{c1}, ds_{c2}, \ldots ds_{cn}$ are, respectively, the central supply to the granule cells and the exterior input to the mitral.

2.2.3 Forward movement towards the target:

Sharks go toward greater odour particles in each area when blood is spilled into the water with a velocity of "V," getting closer to their prey (target). In order to determine the velocity for each dimension,

$$v_{i,1}^{k} = \mu k.R1.\frac{\partial(OF)}{\partial xj} \quad (10)$$

Where k=1,2, ...k_{max} $\frac{\partial(OF)}{\partial xj}$ is the objective function's (OF) derivative at location $x_{(i,1)}^{k}$ is the number of stages, k max is the maximum number of stages for the Shark to move ahead,μk is a value in the range [0, 1], and R1 stands for a random integer in the range [0, 1]. Shark's velocity increases in proportion to the severity of the odour. The Shark's acceleration is constrained by its inertia. Therefore, the speed of the shark now is dependent upon its speed in the past, which may be used by changing equation (10) as illustrated in (11).

$$v_{i,1}^{k} = \mu k.R1.\frac{\partial(OF)}{\partial xj} + a\ k.R2v_{i,1}^{k-1} \quad (11)$$

where [0, 1], $v_{i,1}^{k-1}$ Shark's prior velocity,ak is the inertia coefficient in the range [0, 1], and R2 is a random value in the range [0, 1] much as R1. Shark is moving forward, hence its new location, $Y_{i,1}^{k+1}$, is calculated using its prior position, x_i^k and velocity v_i^k. Consequently, the Shark's new posture might be described as in (12).

$$Y_{i,1}^{k+1} = x_i^k + v_i^k.\Delta t_k \quad (12)$$

where Δt_k is a time period that, for the sake of simplicity, is set to 1.

2.2.4 Rotational movement

Additionally, the Shark rotates in order to detect stronger odour particles. Local search is the term for this SSO algorithm procedure, which may be explained as in (13).

$$Z_{i,1}^{k+1,m} = Y_i^{k+1} + R3.Y_{i,}^{k+1} \quad (13)$$

where m = 1, 2,..., M and R3 is a randomized integer from [1, 1] to M. The rotating movement of the Shark in the search space is modelled in the local search by connecting a large number of locations (M) to create closed contour lines.

2.2.5 Updating the particle position

The Shark's circular movement will continue as it draws closer to the spot and develops a better sense of smell. This SSO algorithm feature may be described as follows:

$$x_i^{k+1} = \arg\max\{OF(Y_i^{k+1}), OF(Z_i^{k+1,i}), \ldots OF(Z_i^{k+1,M})\} \quad (14)$$

where x_i^{k+1} the Shark in the next position with the greatest objective function (OF) value.

2.2.6 Achieving global and local solution

From the equation. $\omega_x(X) = \{f_x(x_1), f_x(x_2) \ldots . f_x(x_n)\}$ and $\omega_y(Y) = \{f_y(y_1), f_y(y_2) \ldots . f_y(y_n)\}$ indicate the outputs of the mitral and granule cells, respectively; f x and f y mimic the cell output functions for the mitral and granule cells; one obtains the mitral outputs at 0.14 and the granule cell outputs at 0.18.

$$f_x(x) = \begin{cases} 0.14 + 0.14\tan h(\frac{x-\varphi}{0.14}) \\ 0.14 + 1.4\tan h(\frac{x-\varphi}{1.4}) \end{cases} \quad (15)$$

$$f_y(y) = \begin{cases} 0.29 + 0.29\tanh(\frac{x-\varphi}{0.14}) \\ 0.29 + 2.9\tanh(\frac{x-\varphi}{2.9}) \end{cases} \quad (16)$$

The synaptic-strength connection matrixes H_0, W_0 and L_0 represent the relationships between mitral cells and granular cells, respectively, and are calculated using the formula as follows.

$$H_{0_i}^{j} = \frac{rand()}{T_h}, 0 < d_i^j < T_h \quad (17)$$

$$W_{0_i}^{j} = \frac{rand()}{T_w}, 0 < d_i^j < T_w \quad (18)$$

$$L_{0_i}^{j} = \frac{rand()}{T_l}, 0 < d_i^j < T_l \quad (19)$$

There are three connection constants: T_h, T_w and T_l. Rand() returns a value at random. The j^{th} odour is the preferred scent for bears, and d_i^j displays the separation between i^{th} and j^{th} smells as a result of their information. To put it another way, this separation is established between each odour (local solution) and the desired odour (global solution). It demonstrates how the optimization process uses the directed mechanism based on global solution to enhance exploitation. The separation process starts based on a dissimilarity evaluation once the brain has gathered all the information from neuronal activity, according to the aforementioned descriptions. The Pearson correlation is the basis for simulating this process. As a result, this point aids in choosing the optimum route for next position. Let the likelihood of the odour components (POC), odour fitness (OF), and likelihood of the odour (POF) defined by:

$$POC_i = \frac{O_i}{\max(O_i)} \quad (20)$$

$$POF_i = \frac{OF_i}{\max(OF)} \quad (21)$$

Expected odour fitness (EOF) and distant odour components (DOC) formulations can be used to determine how distinct two scents are.

$$DOC_i = 1 - \frac{\sum_{i=1}^{k}(POC_j^1 - POC_j^2}{\sqrt{\sum_{i=1}^{k}(POC_j^1 - POC_j^2)^2}} \quad (22)$$

$$EOF_i = (POF_i - POF^g) \quad (23)$$

where g stands for the whole answer. The possible movement is shown by the equations above. These indices actually describe the relationship between scents and how they get to the target place. It is amply demonstrated that the brain's outputs determine the proper route for the following location. The distance between each odour in a mesh grid region is computed using two criteria.

The HSBOA algorithm is employed at this point to extract the best characteristics. First, choose where the Shark and Bear will start from; this point will be in the middle of the data. Secondly, using the fitness function, the fit or good has been discovered for each place near the shark and bear (F). To extract the best characteristics, use the HSBOA algorithm in the third step. The HSBOA method was used in this work to extract 21 characteristics from each set of data after performing 21 iterations. Only one feature with the greatest fitness value is retrieved for each cycle. Fitness value, the placement of the shark and bear during the iteration was changed to either forward or rotating. Shark's position is updated if the fitness value of the site in its forward movement is greater than its fitness value in its

rotating movement. The fitness rating at that point determines whether Shark will go forward or rotate there; also, the areas visited by the HSBOA algorithms cannot be visited again.

2.3 Fusion of multimodal features

Early (or feature-level) fusion combines low-level characteristics from each modality through correlation, which may be more effective but has trouble synchronizing the temporal input from several sources. The final decision is calculated by integrating the unimodal decision values obtained by the late (or decision level) fusion. Let us consider a feature extracted dataset $L = \{(x_1, x_2, \ldots . x_n y_n)\}$ that contains n observations. Each observation $x_n y_n$ contains two elements which are variables, and the predictor is denoted consecutively by $X = (X^1, X^2, \ldots X^p)$ and Y €y where y is either a class label or a numerical response and X€x. The function t: $R^p \rightarrow Y$ is a mapping for a classification problem, where $Y = y$, but is not the same in the regression problem where Y is supposed to be $s(X) + \mu$ where $\mu = 0$ and s indicates the so-called regression function.

2.4 Construction of RNN+Att_ResNet

We will describe in detail our Attention-based Residual Network (Att_ResNet) model using Recurrent Neural Network (RNN), as shown in fig.2, in this part. The formalizations of document-level sentiment categorization are presented first. After that, we go over how to use the Attention-based ResNet, which integrates both forward and backward information, to derive document semantic representation. The input for representing information is the improved sentence representation.

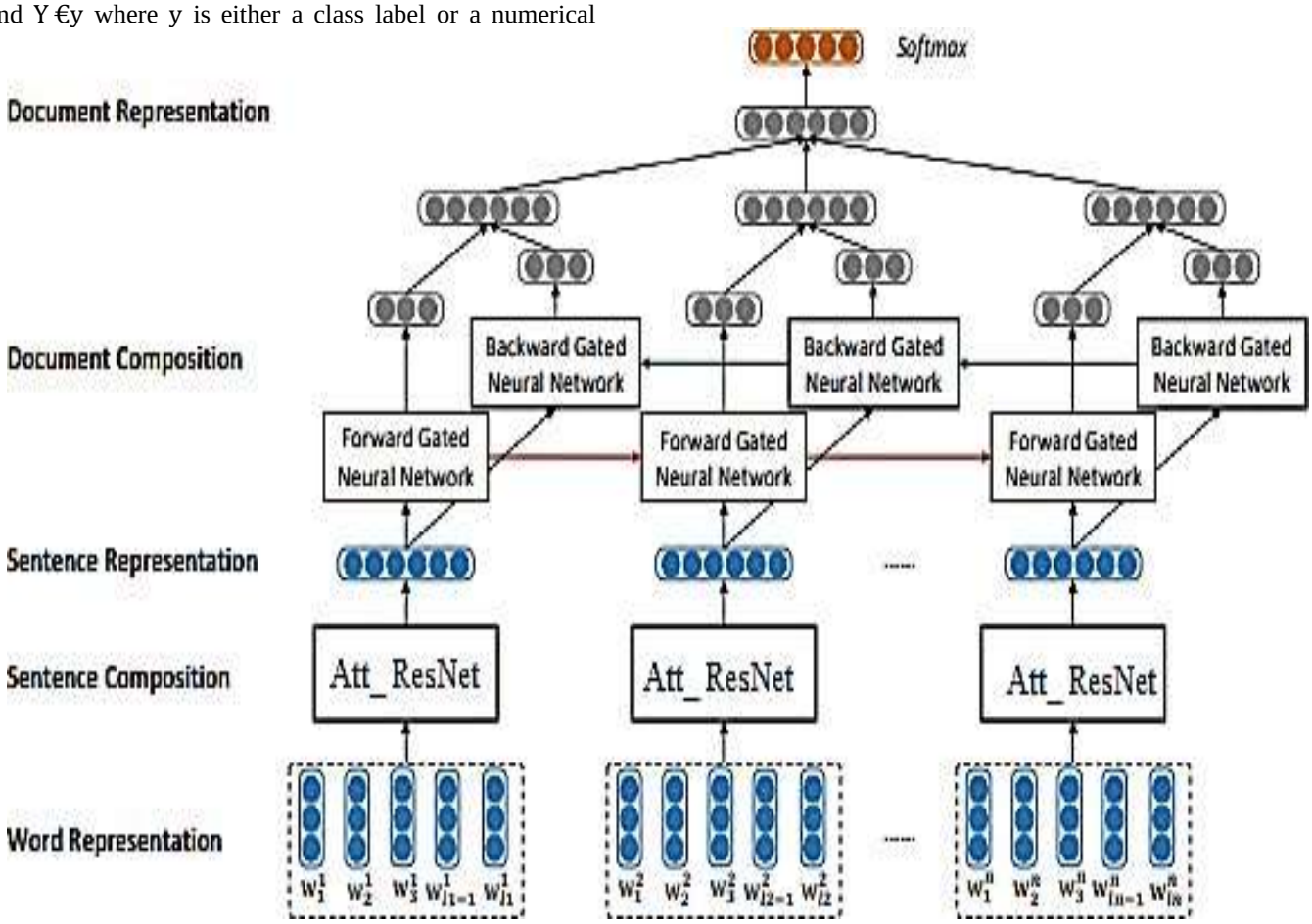

Fig.2 RNN+Att_ResNet based classification process

2.4.1 Formalization

We give the review as a document (d) with the following word count: $\{w_1, w_2, w_3 \ldots w_n\}$. The goal of document-level sentiment classification is to determine the polarity (positive, negative) or numerical rating (one to five or one to ten stars) of these reviews based on the textual content. By iteratively converting the current sentence vector s_t with the output vector of the prior step, h_{t-1} standard recurrent neural network (RNN) can transfer sentences of variable length to sentences with fixed length (t-1). The transition function generally consists of a pointwise nonlinearity layer like tanh, followed by a linear layer.

$$h_t = \tanh(w_r . [h_{t-1}; s_t] + b_r) \quad (24)$$

where $w_r €R^{l(l_h + l_{oc})}$, $b_r €R^l$ and l_{oc} are the hidden vector's and the sentence vector's dimensions, respectively. Particularly, the following formula is utilised to compute the gated RNN's transition function in this study:

$$i_t = sigmoid\ (w_i . [h_{t-1}; s_t] + b_i) \quad (25)$$

$$f_t = sigmoid\ (w_f . [h_{t-1}; s_t] + b_f) \quad (26)$$

$$g_t = tanh\ (w_r . [h_{t-1}; s_t] + b_r) \quad (27)$$

$$h_t = tanh\ (i_t . g_t) + f_t . h_{t-1}) \quad (28)$$

where w_i , w_f , b_i , b_f adaptively choose and delete historical vector and input vector for meaningful composition, where "." denotes element-wise multiplication.

2.4.2 Att_ResNet block

Each word is placed into a lower dimensional semantic space using word2vec. Let W be a word embedding produced by an unsupervised algorithm like Glove or SkipGram, where |V| represents the size of the vocabulary and d is the dimension of the word vectors. Due to the restricted corpus available for word embedding training, embedding could be fine-tuned during the coaching of our framework. The ResNet Block's job is to build the high-level representation by reading the embedding vectors for the input text. Formally, a construction block is defined in this text as

$$y = F(x, w).x \quad (29)$$

The inlet and outlet vectors for the ResNet block are shown here as x and y. Three convolutional layers, each followed by a batch normalisation layer and an activation of the ReLU, make up a ResNet block. The filters and kernel sizes for every convolution are 5 and 200, respectively. The gates' intuition is comparable to the gated recurrent hidden units, in which the gates regulate how much each source contributes to the desired emotion. This is driven by the fact that the left context, the right context, or a mix of both might dominate sentiment signals. We employ gates to govern the combination of h_l and h_r in forming h_{rl}. Formally, in producing h_{rl}, much like the gated recurrent neural network layer, which combines $h_{(i-1)}$ and x_i to form h_i. Formally,

$$h_{rl} = \tan \mathrm{h}\, w_1\, (i_t.\, g_t) + (f_t.\, h_{t-1}) + b_1 \quad (30)$$

The linear interpolation between h_l, h_r and h_{rl}. can be formulated as

$$h_{rl} = z_l.\, h_l + z_r.\, h_r + z_{rl}.\, h_{rl} \quad (31)$$

wherez_l, z_r and z_{rl} are the gate control weights, $z_l + z_r + z_{rl}$r = 1, and

$$z_l \infty \exp\,(w_4 h_l + U_4 h_t + b_4) \quad (32)$$

$$z_l \infty \exp\,(w_5 r + U_5 h_t + b_5) \quad (33)$$

$$z_l \infty \exp\,(w_6 r + U_6 h_t + b_6) \quad (34)$$

In order to obtain h_l, h_t, and h_r, we perform dimension reduction to the pooling outcome of the left contextual, the target, and the right context.

$$h_l = \tanh\,(pooling(left_{context}).\, w_7 + b_7 \quad (35)$$

$$h_t = \tanh\,(pooling(target).\, w_8 + b_8 \quad (36)$$

$$h_r = \tanh\,(pooling(right_{context}).\, w_9 + b_9 \quad (37)$$

In the above equations, the matrices w1, w2, ⋯, w9, u1, u2, ⋯, u6 and the vectors b1, b2, ⋯, b9 are model parameters as shown in the fig. 3.

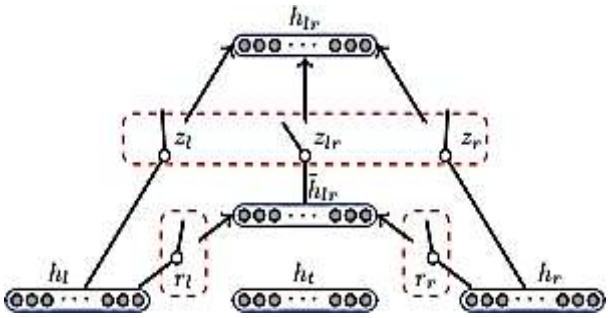

Fig.3 Att_ResNet block representation

2.4.3 Attention block

Although the ResNet Block has a greater capacity, not all words equally contribute to the categorization of mood. The key elements of high-level semantic can be captured with great success by the attention mechanism. The improved sentence representation, in formal terms, is a weighted combination of hidden layers as follows:

$$s_i = \sum_{j=1}^{i1} a^i.\, h^i \quad (38)$$

where a^iindicates how significant the j^{th} word or sentence is for the present sentence. For each hidden state, the attention weight hi may be defined as follows:

$$h^i = \frac{\exp\,(e(h^i))}{\sum_{k=1}^{i} \exp\,(e(h^l))} \quad (39)$$

where e is a scoring function that determines the weight given to words in a document's representation. The definition of the scoring function e is

$$e(h^i) = v^t \tanh\,(w_h h^i + b) \quad (40)$$

where $w_h h^i$ are matrices of weights for hidden state.

The three options for the annotators were positive (+1), neutral (0), and negative (-1). All three modalities were assigned polarities (visual, audio and text). We don't utilize lemmatization or stemming to group related words together or change words to their base form for better representation. We use the word n-gram, which includes unigrams and bigrams, and the character n-gram, which ranges from 2 to 4-g; The relevance of a phrase in each document in the corpus is then highlighted using the tf-idf vectorizer from the scikit-learn python library13. For this, we select and exclude the categories where the scores obtained indicate a little difference between the two groups after normalizing the scores of each category utilizing MinMax normalisation. Tweets from users in the control and diagnosis groups are represented in the equation below, respectively.

$$difference = (D_{neg_{tweets}}) - (D_{pos\ tweets}) \quad (41)$$

$$diff_{score} = \{moderate\_depression \quad (42)$$

$$high\_depression$$

ROUGE-2, the following bigram score (R2). ROUGE-SU4, (SU4), the bigram score that permits up to a 4-word skip.

Score for Bigram coverage (Coverage). This score is comparable to ROUGE-2 but does not take into consideration how often the bigram appears in the human descriptions or the summary that will be assessed. If I out of n human summaries contain a certain bigram, then that

bigram receives a credit of i. ROUGE-2 that is not typical (Bigram). Without the normalization for the length of the summaries, the score is effectively ROUGE-2.
A point-to-point comparison of Bigram coverage (Coverage P2P). This rating is comparable to the third rating.
Measured by a point-to-point comparative, unnormalized ROUGE-2 (Bigram P2P). The point-to-point form of this score.

3. EXPERIMENTAL ANALYSIS

We have tried values of hyperparameters manually for our proposed models and found that those specific values which are mentioned in Table 1 are performing well with our dataset.

Table. 1 Hyper Parameters and their corresponding values

Model	Hyper parameter	Value
Att_ ResNet	Image, video size	224 × 224 × 3
	Function of activation	ReLU, Softmax
	Loss function	Cross correlation
	Optimizer	adam
	Epoch	230
	Batch size	38
RNN	Maximum comment length	35
	Dimension of embedding	160
	Quantity of gated layers	4
	Quantity of pooling layers	2
	Quantity of dense layers	1
	Quantity of filters	3
	Filter size	4
	Pooling size	5
	Dropout rate	3

3.1 Dataset description

We made advantage of a sizable depression dataset that is freely accessible, as suggested by Shen et al. [29]. The writers crawled and annotated the tweets. Three elements make up the dataset: (1) Depressed dataset D1, which includes 2558 samples of users who have been classified as depressed and their tweets, and (2) non-depressed data D2, which includes 5304 samples of users who have been classified as non-depressed and their tweets. (3) The D3 dataset for the Depression Candidate. A sizable, unlabeled dataset of 58810 items representing depression candidates was created by the authors. We employed the D1 and D2 labelled datasets in our investigations. In order to ensure that we have enough statistical data for each user, we pre - process the dataset by excluding users whose posting history includes fewer than 10 posts, users with more than 5000 followers, and users who tweet in a language other than English. Thus, we have taken 4208 users into account (51.30 percent depressed and 48.69 percent non-depressed users). We divided the dataset into training (80%) and test (20%) sets for assessment purposes.

3.2 Selection of size of features from video, image and text

To categories the postings, we tried with various feature sizes for text, photos, and video. We experimented with several feature size combinations, including 256 for images and 128 for text, 256 for images and 256 for text, 256 for videos and 512 for text, 256 for images and 1024 for text, 512 for images and 512 for video, and 256 for images and 1024 for text.

3.2 Performance metrics

Accuracy presents the ability of the overall prediction produced by the model. True positive (TP) and true negative (TN) provides the capability of predicting the absence and presence of negative review. False positive (FP) and false negative (FN) presents the false predictions made by the used model.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (43)$$

Precision- The ratio of a positive sample number is the rate of precision. Precision, on the other hand, refers to the percentage of prediction models in a phrase where the unwanted term is really present.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (44)$$

Recall- The sensitivity calculation does not take into consideration the findings of indeterminate tests because they cannot be replicated, and inconclusive samples should all be omitted from analysis. It is about the detection ability to reliably discover positive reviews in datasets.

$$\text{recall} = \frac{TP}{TP+FN} \quad (45)$$

F1-score – It is used to assess the accuracy of predictions. It represents the weighted average of recall and accuracy. The best value is 1, while the poorest value is 0. F1-score is calculated without taking TNs into account.

$$f1 - score = \frac{2*P*R}{P+R} \quad (46)$$

4. RESULTS

Table.2 Results of utilizing trained deep neural network models to identify depression postings

Method	class	accuracy	precision	recall	F1-score
VGG-16 + CNN	Less_depression	0.76	0.78	0.67	0.76
	Moderate_depression	0.56	0.65	0.57	0.77
	High_depression	0.46	0.65	0.58	0.75
VGG-19 + CNN	Less_depression	0.73	0.58	0.76	0.76
	Moderate_depression	0.56	0.77	0.73	0.67
	High_depression	0.57	0.65	0.72	0.57
Inception + CNN	Less_depression	0.76	0.65	0.71	0.69
	Moderate_depression	0.65	0.68	0.67	0.68
	High_depression	0.76	0.64	0.65	0.65
Xception + CNN	Less_depression	0.76	0.66	0.68	0.64
	Moderate_depression	0.66	0.67	0.69	0.68
	High_depression	0.65	0.68	0.65	0.65
RNN+Att _ResNet	Less_depression	0.79	0.82	0.74	0.73
	Moderate_depression	0.81	0.74	0.77	0.72
	High_depression	0.85	0.75	0.71	0.71

According to Table 2, we discovered that RNN+Att ResNet outperforms the other models in terms of accuracy, precision, recall, and F1-Score for less depression, moderate depression, and high depression.

Table. 3 Results of multi-modal feature combination in RNN+Att_ResNet

Image features size	Text features size	class	accuracy	precision	recall	F1-score
246	145	Less_depression	0.73	0.75	0.61	0.76
		Moderate_depression	0.42	0.64	0.56	0.77
		High_depression	0.42	0.67	0.55	0.75
245	246	Less_depression	0.67	0.68	0.77	0.76
		Moderate_depression	0.54	0.79	0.72	0.67
		High_depression	0.53	0.64	0.73	0.57
256	432	Less_depression	0.76	0.67	0.74	0.69
		Moderate_depression	0.63	0.69	0.68	0.68
		High_depression	0.74	0.63	0.66	0.65
254	1452	Less_depression	0.73	0.63	0.67	0.64
		Moderate_depression	0.60	0.65	0.65	0.68
		High_depression	0.65	0.68	0.67	0.65
512	512	Less_depression	0.73	0.81	0.74	0.73
		Moderate_depression	0.76	0.77	0.71	0.72
		High_depression	0.79	0.75	0.76	0.71

As shown in Table 3, we discovered that, in high accuracy, precision, recall, and F1-Score for less depression, moderate depression, and high depression, the characteristic of size 512 for both picture and text performs better than other feature combinations.

Table. 4 Prediction Results for decision level fusion

Modality		accuracy	precision	recall	F1-score
Audio+video	positive	0.82	0.89	0.87	0.86
	negative	0.84	0.97	0.89	0.89
Video+text	positive	0.89	0.86	0.85	0.85
	negative	0.87	0.83	0.86	0.89
Audio+text	positive	0.87	0.89	0.87	0.95
	negative	0.85	0.87	0.91	0.94
Audio+video +text	positive	0.94	0.94	0.84	0.92
	negative	0.88	0.85	0.86	0.84
Text only	positive	0.91	0.88	0.97	0.87
	negative	0.93	0.91	0.88	0.91
Audio only	positive	0.86	0.84	0.89	0.93
	negative	0.86	0.96	0.97	0.85

Video only	positive	0.87	0.87	0.86	0.84
	negative	0.8	0.87	0.95	0.83

Table 4 depicts decision level fusion prediction results.

Table. 5 Prediction Results: Text-Based multimodal sentiment analysis

	accuracy	precision	recall	F1-score
positive	0.92	0.83	0.87	0.87
negative	0.88	0.94	0.91	0.86
neutral	0.93	0.92	0.89	0.92
average	0.91	0.93	0.91	0.90

Table 5 depicts text based multimodal sentiment analysis prediction results.

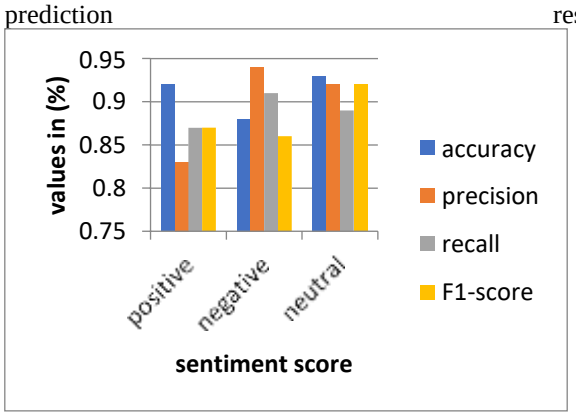

Fig.4 Text based multimodal RNN+Att_ResNet classification

Fig. 4 shows a comparison of recall, precision, and accuracy and F1-score with respect to positive, negative and neutral sentiment classes for classifying the text feature using the proposed RNN+Att_ResNet where X axis shows the sentiment score used for analysis and Y axis shows the values obtained in percentage. When compared, the text analysis achieves 91% of accuracy, 93% of precision, 91% of recall and 90% of F1-score.

Table. 6 Prediction Results: audio-Based multimodal sentiment analysis

	accuracy	precision	recall	F1-score
positive	0.94	0.84	0.85	0.86
negative	0.86	0.92	0.93	0.84
neutral	0.96	0.91	0.92	0.91
average	0.92	0.91	0.93	0.92

Table 6 depicts audio based multimodal system analysis prediction results.

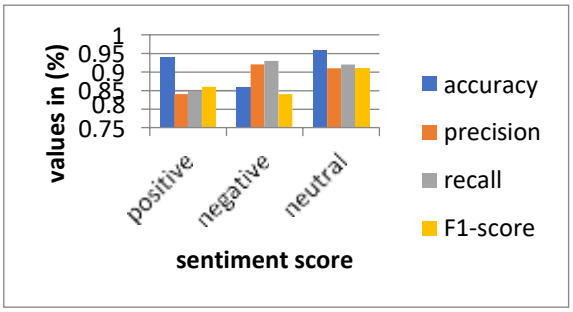

Fig.5 Audio based multimodal RNN+Att_ResNet classification

The fig. 5 illustrates the comparison of accuracy, precision, recall and F1-score with respect to positive, negative and neutral sentiment classes for classifying the audio feature using the proposed RNN+Att_ResNet where X axis shows the sentiment score used for analysis and Y axis shows the values obtained in percentage. When compared, the text analysis achieves 92% of accuracy, 91% of precision, 93% of recall and 92% of F1-score.

Table 7: Prediction Results: video-Based multimodal sentiment analysis

	accuracy	precision	recall	F1-score
positive	0.90	0.83	0.83	0.83
negative	0.88	0.91	0.92	0.84
neutral	0.95	0.94	0.91	0.83
average	0.93	0.83	0.92	0.82

Table 7 depicts video based multimodal sentiment analysis prediction results.

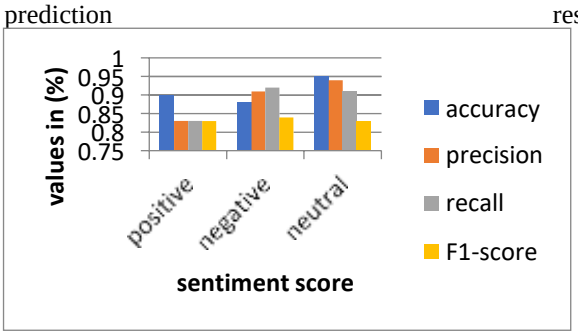

Fig.6 Video based multimodal RNN+Att_ResNet classification

The fig.6 illustrates the comparison of accuracy, precision, recall and F1-score with respect to positive, negative and neutral sentiment classes for classifying the video feature using the proposed RNN+Att_ResNet where X axis shows the sentiment score used for analysis and Y axis shows the values obtained in percentage. When compared, the text analysis achieves 93% of accuracy, 83% of precision, 92% of recall and 82% of F1-score.

Table.8 overall comparative analysis for multimodal features

modality	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
text	91	93	91	90
audio	92	91	93	92
video	93	83	92	82

Table 8 depicts overall comparative analysis of multimodal characteristics.

5. CONCLUSION

This study uses self-report diagnosis to find depressed Twitter users. By initially attempting to forecast mental disorders via social data mining, we then attempted to create a novel framework by automatically gathering information from the profiles of the diagnostic and control users. We used the pre-processing techniques and retrieved the required characteristics from the dataset. To determine the differences between the two groups and confirm earlier findings, we next looked at the distribution of data in predetermined categories. Additionally, automatic text summarization has the advantage that there are no arbitrary design decisions driving our feature selection, such as excluding sentences that contain specific preset terms or phrases that are within a specific length. In comparison to the competing models, our RNN+Att ResNet model can learn more discriminative multi-modal characteristics from data. We get noticeably higher outcomes than current state-of-the-art baselines thanks to the interaction of automated summarization, user behavioral representation, and model training. In the future, we'll employ URLs, photos, a combination of short and long user-generated material, as well as standard web sites in addition to social networking content. This would assist provide the model with more contextual information and allow us to concentrate on a task wherein our system not only recognizes sadness but also automatically offers potential diagnoses.

REFERENCES

1. A.Bogomolov, B. Lepri, M.Ferron, F. Pianesi & A. Pentland(2014), Daily stress recognition from mobile phone data, weather conditions and individual traits, Proc. of the 22nd ACM international conference on Multimedia (pp.477-486). https://doi.org/10.1145/2647868.2654933
2. C. Buckley & E.M. Voorhees(2004), Retrieval evaluation with incomplete information, Proc. of the 27th annual international ACM SIGIR conference on Research and development in information retrieval (pp. 25-32). https://doi.org/10.1145/1008992.1009000
3. X. Chang, Y. Yang, A. Hauptmann, E.P. Xing & Y.L. Yu (2015), Semantic concept discovery for large-scale zero-shot event detection, Twenty-fourth international joint conference on artificial intelligence.doi:https://www.ijcai.org/Proceedings/15/Papers/316.pdf
4. W. Che, Z. Li & T. Liu (2010), Ltp: A chinese language technology platform, Coling 2010: demonstrations (pp. 13-16).doi: https://aclanthology.org/C10-3004/
5. C. C. Chang and C.-J. Lin.(2001) ,Libsvm: A library for support vector machines, ACM Trans. Intell. Syst. Technol. vol. 2, no. 3 (pp. 389–396). https://doi.org/10.1145/1961189.1961199
6. D. C. Ciresan, U. Meier, J. Masci, L. M. Gambardella, and J. Schmidhuber(2011),Flexible, high performance convolutional neural networks for image classification,Proc. Int. Joint Conf. Artif. Intell (pp. 1237to1242).doi:https://people.idsia.ch/~juergen/ijcai2011.pdf
7. G. Coppersmith, C. Harman, and M. Dredze.(2014),Measuring post traumatic stress disorder in twitter,Proc. Int. Conf. Weblogs Soc. Media(pp. 579–582). https://doi.org/10.1609/icwsm.v8i1.14574
8. F. A. Pozzi, D. Maccagnola, E. Fersini, and E. Messina(2013),Enhance user-level sentiment analysis on microblogs with approval relations, Proc. 13th Int. Conf. AI* IA: Advances Artif. Intell(pp. 133–144).doi: 10.1007/978-3-319-03524-6_12
9. X. Wang, J. Jia, P. Hu, S. Wu, J. Tang, and L. Cai.(2012),Understanding the emotional impact of images, Proc. 20th ACM Int. Conf. Multimedia(pp. 1369–1370).doi: 10.1145/2393347.2396489
10. C. Lee & J.F. Coughlin (2015),PERSPECTIVE: Older adults' adoption of technology: an integrated approach to identifying determinants and barriers, Journal of Product Innovation Management, 32(5),747-759.doi: https://doi.org/10.1111/jpim.12176
11. T. S. Reddy, Y. D. S. Raju (2021), “Implementation of Data Security with Wallace Tree Approach Using Elliptical Curve Cryptography on FPGA,” Turkish Journal of Computer and Mathematics Education, Vol. 12, pp. 1546-1553. Doi:10.17762/turcomat.v12i6.2693
12. S. M. Schueller, M. Neary, K. O’Loughlin, and E. C. Adkins.(2018),Discovery of and interest in health apps among those with mental health needs: Survey and focus group study. J. Med. Internet Res, vol. 20, no. 6.
13. H. Zogan, I. Razzak, S. Jameel, & G. Xu(2021), Depressionnet: A novel summarization boosted deep framework for depression detection on social media, arXiv preprint arXiv:2105.10878.doi: https://doi.org/10.48550/arXiv.2105.10878
14. Y. Ding, X. Chen, Q. Fu, & S. Zhong (2020), A depression recognition method for college students using deep integrated support vector algorithm, IEEE access, 8,75616-75629. doi: 10.1109/ACCESS.2020.2987523
15. N. Kaushik & M.K. Bhatia (2020),Information retrieval from search engine using particle swarm

optimization, Advances in Computing and Intelligent Systems: Proceedings of ICACM 2019 (pp. 127-140), Springer Singapore.doi: 10.1007/978-981-15-0222-4_11

16. N. Kaushik & M.K. Bhatia (2022),Study of E-bay Used Cars Sales using Tableau, In Jagannath University Research Journal (JURJ), ISSN: 2582-6263, Volume No.-II, Issue No.–II.doi: https://www.jagannathuniversity.org/assets/jnu-docs/jurj/publication-paper/1648189874-jurj221119.pdf
17. A. Chugh, V.K. Sharma, S. Kumar,A. Nayyar, B. Qureshi,M.K. Bhatia & C. Jain(2021), Spider monkey crow optimization algorithm with deep learning for sentiment classification and information retrieval, IEEE Access, 9, 24249-24262.doi: 10.1109/ACCESS.2021.3055507
18. N. Kaushik & M.K. Bhatia (2022), Twitter sentiment analysis using K-means and hierarchical clustering on COVID pandemic, International Conference on Innovative Computing and Communications: Proceedings of ICICC 2021, Volume 1 (pp. 757-769). Springer Singapore.doi: https://doi.org/10.1007/978-981-16-2594-7_61
19. A. Chugh, V.K. Sharma, M.K. Bhatia & C. Jain. (2022),A big data query optimization framework for telecom customer churn analysis, International Conference on Innovative Computing and Communications: Proceedings of ICICC 2021, Volume 2 (pp. 475-484), Springer Singapore.doi: 10.1007/978-981-16-2597-8_40
20. N. Kaushik, M.K. Bhatia, & S. Rastogi (2020), Various data classification technique on crime dataset. International Journal of Management, Technology and Engineering, Pp. (84-92).

AUTHORS

Nainika Kaushik has received her bachelor's degree in computer science and engineering from GGSIPU, India in 2013 and PG in mobile and pervasive computing (computer science engineering) from IGDTUW, India in 2015. She is currently pursuing PhD at the Department of Computer Science and Engineering, Jagannath University, Jaipur, India. Her areas of interest are data mining and machine learning.

Corresponding Author **E-mail:** nainika.kaushik@jimsindia.org

Manjot Kaur Bhatia has received her bachelor's degree in computer applications from GNDU, India in 1994 and PG in computer applications from IGNOU, India in 2002. Ph.D. in computer science from Delhi University, India in 2016.Her area of interest includes Cloud Computing, Steganography, Data Hiding and Information security.

E-mail: manjot.bhatia@jimsindia.org