



# Mean Silhouette Genetic and Elman Deep Recurrent Network based Sentimental Analysis for Twitter Social Media Healthcare Data

<sup>1</sup>Dr.D.Sasikala,  
Assistant Professor  
Department of Computer Science,  
Sri Vasavi College,  
Erode.

<sup>2</sup>Dr.S.Sukumaran  
Associate Professor  
Department of Computer Science,  
Erode Arts and Science College  
Erode.

sasikalad@gmail.com

## Abstract

Employment of social media is becoming all-pervasive and disease analogous communities are organizing online, together with communities of attentiveness encompassing health care domain. Despite Facebook being the most sought out social media platform, utilization of supplementary social media platform like Twitter is increasing. To be more specific, in recent days patients with COVID and diabetes commenced to gather and take active participations in online discussions about diabetes and COVID on Twitter, engross in communication and sharing virtually and perceive peer support online. In this work a method called, Mean Silhouette-based Genetic and Elman Deep Sentiment Analysis (MSG-EDSA) is proposed. Elman Deep Recurrent Network Sentiment Classification is applied to the selected features, to classify the tweets in an accurate and timely manner. The tweets are finally classified as extremely positive, positive, extremely negative, negative or neutral from the tweets obtained via different users. The proposed MSG-EDSA is experimented with using the diabetes and COVID real-time datasets from social media platforms to analyze healthcare sentimental analysis. The parameters like, precision, recall, accuracy and error rate are selected to analyze the performance against the state-of-the-art sentimental analysis methods.

2340

**Keywords:** Twitter, Sentimental Analysis, Mean Silhouette, Genetic, Feature Selection, Elman, Deep Recurrent Network, Classification

**DOI Number:** 10.48047/NQ.2022.20.20.NQ109235

**NeuroQuantology2022;20(20): 2340-2355**

## 1. INTRODUCTION

Recognizing the presence or occurrences of emotions in text sequence is usually contemplated to be a laborious and cumbersome chose owing to the long term

dependency in a text. In the recent few years, there has been an enormous design and development in the area of emotion analysis however analysis concerning text sequence remains to be a paramount task to apprehend



about the emotion patterns produced by sequential text. Numerous emotion analysis models have been designed however with the inclusion of text-related requests along with depression measures have to be taken to address specific type of depression. Multi Head Attention with Bidirectional long short-term memory and Convolution Neural Network (MHA-BCNN) was proposed in [1] for acquiring the negative text-based emotions. Here, by extracting deep learning model, numerous negative mental-health emotions including, addiction, depression, stress, insomnia, stress, long term dependencies were analyzed. Also, by applying the GloVe embeddings, ambiguity factors were handled in a definite progression, therefore improving the precision and recall with maximum accuracy.

Though precision and recall with maximum accuracy was attained, the error rate involved in the emotion analysis was not focused. To address this issue, in our work, Mean Silhouette-based Genetic Feature Selection model is applied to the pre-processed tweets that with the aid of Mean Silhouette in the Genetic model, selects the robust and relevant features or attributes with minimum error. Despite of the victory of neural networks mechanisms in enhancing learning structural responsibility, they emanate insufficient precision in bio-medical sentiment classification when utilizing rarely functional or convenient features sets. A deep associative memory into neural networks was proposed in [2] with the purpose of decomposing sentiment that in turn spotlighted precisely on bio-medical structures concerning to the data object extraction for a specified characteristic. With the deep associative function, unseen medical patterns were efficiently measured, ensuring accuracy. Despite accuracy ensured, the other parameters like precision and recall was not focused. To address this aspect, Elman Deep Recurrent Network Sentiment Classification model is applied to the selected features that with the aid of context unit in deep learning

eISSN1303-5150

ensure both recall and precision in a timely manner.

This work is to analyze the sentiments on the basis of the tweets posted by the users in Twitter and rank the polarity score on diabetes and corona virus tweet dataset to understand users' opinions and to discover facts by analyzing the positive and negative tweets in healthcare domain. This study concentrates on a given time point along with acquiring the sentiments and makes an effort to select the relevant feature which may get great attention of the users or patients' community based on the tweet polarities.

The paper is organized as follows: Related works is described in section 2, followed by section 3 explaining our overall Mean Silhouette-based Genetic and Elman Deep Sentiment Analysis (MSG-EDSA) method and feature selection and sentimental analysis classification model. The experimental results on both feature selection and sentimental analysis classification are covered in section 4. In Section 5 discusses in detail regarding the accuracy, precision, recall and error rate concerning sentimental analysis. Section 6 covers the conclusion.

## 2. RELATED WORKS

Over the past few years, as numerous users' opinions are obtainable on the web, sentiment analysis has become one of the most generative research domains as far as Natural Language Processing (NLP) is concerned. Owing to this research on sentiment analysis has focused on an extensive domain like, economy, politics, medicine, and so on. Hence, materials and methods for sentiment analysis have advanced from simple rule applications to sophisticated machine learning mechanisms, like, deep learning. The COVID-19 pandemic has dispensed numerous distinct ultimatums to patient care specifically in catastrophe medicine. These issue consequence in an improved patient experiences. In [3], three

www.neuroquantology.com



periods of time, like, pre, during and post-COVID-19 first wave were covered with the purpose of identifying the concentric tendencies in sentiment and topics related to patient experience. However, very few endeavors have been made to appertain deep learning to sentiment analysis for reviewing numerous drug. A standard comparison of numerous deep learning frameworks like, Convolutional Neural Networks (CNN) and long short-term memory (LSTM) recurrent neural networks were proposed in [4]. Best results were obtained but with a very high training time. Sentiment analysis for diabetes in twitter was analyzed in [5].

Although Twitter messages are frequently utilized in acquiring sentiment scores that are further associated to market performance, particular emotional aspects has not been used in the previous work to analyze risk behavior. In [6], valence, arousal and dominance were associated on the cryptocurrency market with the purpose of predicting future risk. In [7], machine learning tools were applied for analyzing the diabetes acquired through tweets by various patients. Text mining with sentiment analysis was applied in [8] using naïve bayes for analyzing patient records. People agonizing from chronic diseases like diabetes should have cyclic communication with for self management of their condition. Hence, social networks like Twitter is an exceptional supply for the patients as they can bridge with people who have a homogeneous situation and similar experiences. It bestows the environment and the mechanisms for sharing valuable knowledge and peer support. An aspect level sentiment analysis method was proposed in [9] with the corpus obtained from Twitter therefore resulting in the improvement of both precision and recall. Tweets and features associated with obesity and diabetes were analyzed in [10] that in turn contributed to both utility and cost effectiveness on individual health. Yet another Constant Rebalancing and Modern  
eISSN1303-5150

Portfolio Theory was applied in [11] for hybrid portfolio optimization.

The world is noticing an eruption of social networks (SNS) like Twitter where people perform information exchanges, opinions and ideas. Enormous voluminous data acquired on such SNS with the aid of NLP derive meaningful information. One of the most popular techniques of NLP is Sentiment analysis (SA). Sentiments generated during COVID 19 outbreak on twitter was analyzed in [12]. A holistic review on diabetes prediction collected from twitter was investigated in [13]. Based on the tweet results, classification of patients was made by employing machine learning techniques [14] therefore contributing to high precision and recall. An elaborative sentimental analysis on the consequence of corona virus using the Bidirectional Encoder Representations from Transformers (BERT) model was proposed in [15]. With this BERT model, validation accuracy was ensured. Yet another social network analysis with respect to knowledge sharing was investigated in [16]. Both qualitative and quantitative analysis with continual treatment to diabetes with tweets was analyzed in [17]. Semi structured interviews were conducted in [18] promoting patients' self management of diabetes as a means of novel sharing platform. Outliers involved in the self management of diabetes were concentrated in [19]. A survey of deep learning for sentimental analysis was investigated in [20].

2342

### 3. METHODOLOGY

Our proposed Mean Silhouette-based Genetic and Elman Deep Sentiment Analysis (MSG-EDSA) method as illustrated in figure 1 comprises of three modules and integrates with Twitter as a plug-in application.

In the first module pre-processing, analyzing the tweets acquired from the corona virus tweets and diabetes dataset. Next, the pre-processed tweets are analyzed for selecting pertinent and robust subset of

www.neuroquantology.com



features by means of Mean Silhouette-based Genetic Feature Selection model. Finally, with the robust subset of selected features, precisely classifies the tweets of user into extremely positive, positive, extremely negative, negative or neutral using Elman Deep Recurrent Network Classification.

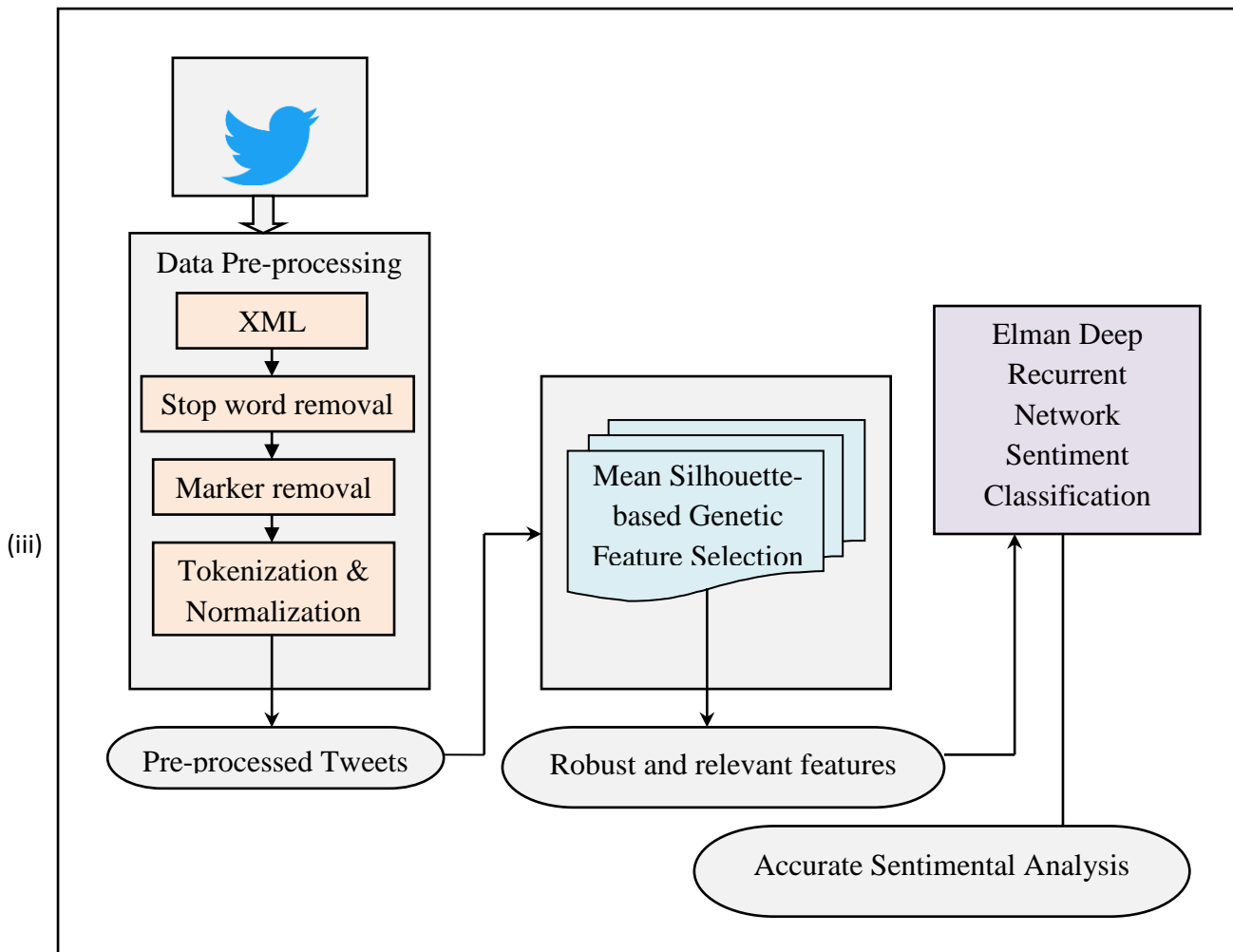
**3.1 Pre-processing**

To attain good results, the first and the foremost process in classifying tweets ‘T’ for diabetic and corona virus related search terms is pre-processing. This is because only after robust analysis of tweets good results are said to be achieved. Numerous distinct

steps are said to be involved in the process of pre-processing. They are

- (i) Removal of URL, Punctuation, User name and stop word, to acquire a standard dataset suitable for further processing.
- (ii) Second the dataset comprises of tweets comprising of numerous HTML tags, punctuation, and multiple spaces that provides no meaning for feature extraction, therefore removals of these markers ensures in the minimization of the noise, therefore enhancing the classification of emotions, hence assisting in real time diabetic sentiment analysis.

2343



**Table 1: Examples of Pre-processed tweets before and after Pre-processing**

Dataset	# Tweets (before pre-	# Tweets (after pre-processing)
---------	-----------------------	---------------------------------



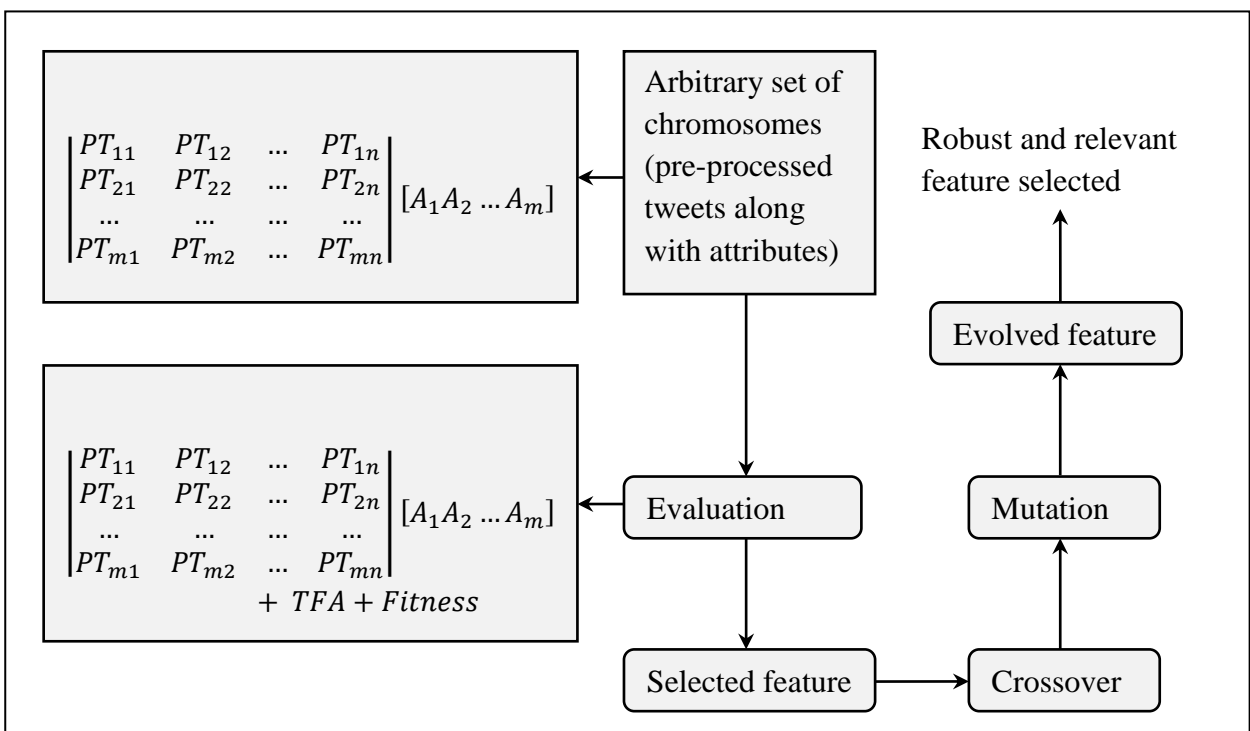
	processing)	
Diabetes dataset	Severe hypoglycemia (Low blood sugar) can even cause seizures, comas and hypothermia!	'Severe' 'hypoglycemia' 'Low' 'blood' 'sugar' 'even' 'seizures' 'comas' 'hypothermia'
Corona Virus Tweets	Studies show the #coronavirus like #COVID19 can live up to nine days on hard surfaces like metal...	'coronavirus', 'COVID19', 'nine days', 'hard surfaces', 'metal'...

### 3.2 Mean Silhouette-based Genetic Feature Selection

Upon successful completion of the pre-processing, the next step for diabetic and corona virus sentiment analysis is feature selection. To be more specific, selection of pertinent or significant tweets from the pre-processed tweet along with the relevant

feature or attribute is referred to as feature selection. This is because to perform sentimental analysis not only the relevant tweets as has to be selected but also the features or attributes so that a precise classification concerning the tweets can be made in an efficient manner.

2344



The range 'r' of each chromosome, with 'r < m' refers to the range of the feature set to be selected from the pre-processed tweets. Then, a chromosome population is

represented in the form of a matrix 'r \* k' with 'k' denoting the chromosome number and the initial population to be created in an arbitrary fashion. Then, the feature selection



problem is formulated as the following optimization problem as given below.

$$\begin{cases} f(PT) \\ \max \text{Such that } PT = (PT_1, PT_2, PT_3, \dots, PT_n), PT_i \in \{0,1\} \\ 1 \leq PT_i \leq n \end{cases} \quad (1)$$

A discrete and differentiating feature of the Mean Silhouette-based Genetic Feature Selection model is the mechanism to assess the fitness of each chromosome (i.e., pre-processed tweets). This is performed via dual succession steps. First, a Tweet Factor Analysis (TFA) of the examined samples (i.e. pre-processed tweets) is carried out utilizing the chromosome attributes (i.e., pre-

$$PT_{i,j} - MV_i = L_{i,j}A_{j,m} + \dots + L_{i,k}A_{k,m} + \epsilon_{i,m} \quad (2)$$

From the above equation (2), Tweet Factor Analysis (TFA) of the examined samples is performed by means of the pre-processed tweets ' $PT_{i,j}$ ' of the ' $i - th$ ' observation of the ' $j - th$ ' user, mean variance ' $MV_i$ ' for the ' $i - th$ ' observation, loading matrix ' $L_{i,j}$ ' of the ' $i - th$ ' observation of the ' $j - th$ ' user,

processed tweet attributes). The objective behind the utilization of TFA is that the pre-processed tweets gained about the interdependencies between observed attributes can be used at a later stage to minimize the set of variables or attributes in a dataset. This is mathematically expressed as given below.

common attributes ' $A_{j,m}$ ' of the ' $j - th$ ' observation of the ' $j - th$ ' user and a negligible error ' $\epsilon_{i,m}$ ' respectively. Then the Mean Silhouette Index (MSI) is evaluated on the sample factors obtained by TFA.

2345

$$MSI = \frac{\sum_{i=1}^n \frac{MM(i) - MV(i)}{\max\{MV(i), MM(i)\}}}{n} \quad (3)$$

From the above equation (3), the Mean Silhouette Index ' $MSI$ ' is obtained based on the mean variance ' $MV(i)$ ' of ' $i$ ' with respect to all other pre-processed tweet samples within the same class (i.e., feature set),

moderate mean ' $MM(i)$ ' distance of ' $i$ ' to all pre-processed tweet samples to other class (i.e., feature set) respectively. Finally, the fitness function is mathematically formulated as given below.

$$Fitness = \begin{cases} MSI + \frac{n}{M}, MSI > 0 \\ 0, MSI + \frac{n}{M} \leq 0 \end{cases} \quad (4)$$

In this manner, with maximum fitness score for each attribute is equal to 1 is selected as the relevant feature, (i.e., the score that are allocated to a chromosome or attributes with the maximum differentiation potentiality than

other attributes), whereas with fitness score for each attribute is equal to 0 is discarded and not selected for further processing or analysis. The selection of chromosome or the attributes with highest fitness score is utilized.



Followed by which, crossover between attributes are performed to obtain their correlations. At the end of this process, robust and relevant features are selected with

minimum error rate. The pseudo code representation of Mean Silhouette-based Genetic Feature Selection is given below.

<b>Input:</b> Tweets Dataset, Tweets ' $T$ ', Tweet log ' $TL$ ', Patients ' $P_i$ ', time instance ' $t_i$ ', Attributes ' $A$ '
<b>Output:</b> Robust and relevant feature subset ' $FS$ '
<p>Step 1: <b>Initialize</b> time instance '<math>t_i</math>', pre-processed tweets '<math>PT</math>'</p> <p>Step 2: <b>Initialize</b> Attributes '<math>A</math>', patients '<math>P_i</math>'</p> <p>Step 3: <b>Begin</b></p> <p>Step 4: <b>For</b> each Tweets Dataset as input</p> <p>Step 5: Formulate optimization problem for feature selection problem</p> <p>Step 6: Evaluate Tweet Factor Analysis '<math>TFA</math>' for acquiring interdependency</p> <p>Step 7: Evaluate Mean Silhouette Index '<math>MSI</math>' on the sample factors obtained by tweet factor analysis</p> <p>Step 8: Evaluate fitness function based on maximum fitness score</p> <p>Step 9: <b>Return</b> relevant features (i.e., attributes) selected '<math>FS</math>'</p> <p>Step 10: <b>End for</b></p> <p>Step 11: <b>End</b></p>

2346

### Algorithm 1: Mean Silhouette-based Genetic Feature Selection

As given in the above Mean Silhouette-based Genetic Feature Selection with the objective of reducing the error involved in the selection process and therefore the classifications of labels, first, pre-processed tweets are obtained as input. Second, tweet factor analysis is evaluated to evaluate the variance between observed, correlated attributes in terms of a probably lower number of unobserved attributes referred to as factors. Followed by which Mean Silhouette Index is estimated on the sample pre-processed

tweets and the attributes. Finally fitness function is measured to obtain the robust and relevant features or attributes with minimum error rate.

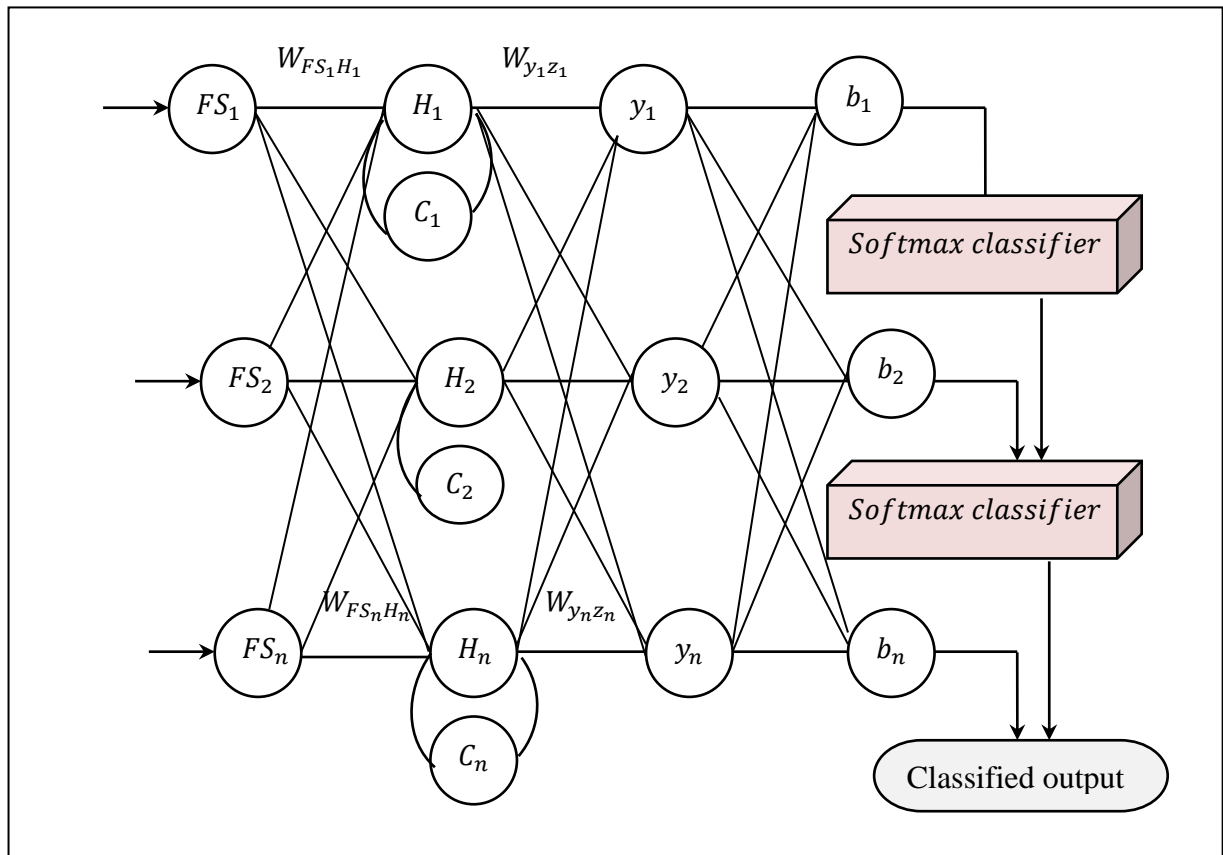
### 3.3. Elman Deep Recurrent Network Sentiment Classification

With the relevant and robust feature selection, classification is performed for sentimental analysis. Sentiment classification refers to the procedure of identifying opinions in tweets and labeling them as extremely



positive, positive, extremely negative, negative or neutral, based on the tweets users express within them. With the assistance of deep learning for sentiment classification emotions can be identified in huge volumes of tweets in a timely and accurate manner. In this work, Elman Deep Recurrent Network Sentiment Classification

model is employed for precise classification of tweets and enhancing sentiment accuracy. Figure 3 given below shows the structure of Elman Deep Recurrent Network Sentiment Classification model.



2347

**Figure 3: Structure of Elman Deep Recurrent Network Sentiment Classification**

As illustrated in the above Elman Deep Recurrent Network Sentiment Classification model, three layers are present, they are, input layer, hidden layer and the output layer in addition to a context unit. The features selection from the Mean Silhouette-based Genetic Feature Selection algorithm forms as input and is modeled in the matrix form, as

' $FS_t = (FS_1, FS_2, \dots, FS_n)$ ' respectively. The hidden layer is connected to the context unit which in turn maintains the previous values of the hidden layers in this unit for future reference. The hidden layer comprises of hidden units modeled as ' $H_t = (H_1, H_2, \dots, H_n)$ ' and is represented as given below.





$$H_t = \sigma_H(W_H FS_t + C_H H_{t-1} + B_H) \quad (5)$$

Finally, the output layer is associated with the hidden units via weights and is mathematically represented as given below.

$$y_t = \sigma_y(W_y H_t + B_y) \quad (6)$$

From the above equations (5) and (6), ' $\sigma_H$ ', ' $\sigma_y$ ' represents the activation function for the hidden layer, output layer, ' $W_H$ ', ' $W_y$ ' representing the weight of the hidden layer, output layer, with bias of the hidden and

output layer denoted as ' $B_H$ ', ' $B_y$ ' respectively. Then, the softmax classifier in the classification of tweets is mathematically formulated as given below.

$$H_A(FS^{(i)}) = \begin{bmatrix} Prob(y^{(i)} = 1 | FS^{(i)}; A) \\ Prob(y^{(i)} = 2 | FS^{(i)}; A) \\ \dots \\ Prob(y^{(i)} = n | FS^{(i)}; A) \end{bmatrix} = \frac{1}{\sum_{i=1}^n e_i^{A_j FS^{(i)}}} \begin{bmatrix} e^{A_1 FS^{(i)}} \\ e^{A_2 FS^{(i)}} \\ \dots \\ e^{A_n FS^{(i)}} \end{bmatrix} \quad (7)$$

2348

From the above equation (7), each sub-vector of vector ' $H_A(FS^{(i)})$ ' denotes the probability value that ' $FS$ ' belongs to different classes or labels (i.e., extremely positive, positive, extremely negative, negative or neutral). Moreover, ' $A_1, A_2, \dots, A_n$ ' denotes the attribute vectors, respectively. After getting the attributes ' $A$ ', the previously resultant values in the form of context unit is utilized and the values or tweets in the hidden layer is processed to estimate the probability value that tweets of the corresponding attribute belonging to each class. The class that possess the highly probability value represents the final classified result.

As given in the below Elman Deep Recurrent Network Sentiment Classification process, the objective remains in designing a classification model with maximum accuracy and in a timely manner. With this objective, an Elman three layer network in a Deep Recurrent model is designed that with the aid of context unit in addition to the hidden layer maintains the previously stored tweets for further process in a timely manner. In addition, softmax classifier in the classification of tweets classifies the tweets and produces the labels in an accurate manner. The pseudo code representation of Elman Deep Recurrent Network Sentiment Classification is given below.

<b>Input:</b> Tweets Dataset Tweets ' $T$ ', Tweet log ' $TL$ ', Patients ' $P_i$ ', time instance ' $t_i$ ', Attributes ' $A$ '
<b>Output:</b> Precise and accurate sentiment classification
Step 1: <b>Initialize</b> features (i.e., attributes) selected ' $FS$ ' Step 2: <b>Begin</b> Step 3: <b>For</b> each Corona virus tweets Dataset ' $CDS$ ' and Diabetes dataset ' $DDS$ ' as input



Formulate the hidden layer with the features obtained from Mean Silhouette-based Genetic Feature Selection

Step 5: Formulate the output layer using weight and bias factors

Step 6: Formulate the softmax classifier for obtaining precise and accurate tweets

Step 7: **If** ' $H_A(FS^{(i)})$  lies between 0 and 0.25'

Step 8: **Then** label is negative

Step 9: **End if**

Step 10: **If** ' $H_A(FS^{(i)})$  lies between 0.25 and 0.50'

Step 11: **Then** label is extremely negative

Step 12: **End if**

Step 13: **If** ' $H_A(FS^{(i)})$  lies between 0.50 and 0.75'

Step 14: **Then** label is positive

Step 15: **End if**

Step 16: **If** ' $H_A(FS^{(i)})$  lies between 0.75 and 1'

Step 17: **Then** label is extremely positive

Step 18: **End if**

Step 19: **If** ' $H_A(FS^{(i)})$  equals 1'

Step 20: **Then** label is neutral

Step 21: **End if**

Step 22: **End for**

Step 23: **End**

2349

**Algorithm 2: Elman Deep Recurrent Network Sentiment Classification**

**4. EXPERIMENTAL SETUP**

The proposed work is tested in the healthcare domain. Experimental evaluations will be performed in Python by utilizing Pima Indians Diabetes Database (<https://www.kaggle.com/uciml/pima->

indians-diabetes-database) and coronavirus tweets NLP text classification (<https://www.kaggle.com/datatattle/covid-19-nlp-text-classification>) dataset. Simulations will be conducted on various metrics like, accuracy, precision, recall and error rate.

**Table 2: Classes of Corona Virus Dataset**

Class	# of Comments	Percentage
Extremely positive	6624	12.3
Positive	18453	34.3
Extremely Negative	5481	10.02
Negative	15509	28.8
Neutral	7718	14.39
Total	53,785	100

**Table 3: Classes of Diabetes Dataset**

Class	# of Comments	Percentage
Positive	268	35
Negative	500	65
Total	768	100



## 5. DISCUSSION

Comparative analysis of sentimental analysis methods is performed and compared with three distinct methods Mean Silhouette-based Genetic and Elman Deep Sentiment Analysis (MSG-EDSA), Multi Head Attention with Bidirectional long short-term memory and Convolutional Neural Network (MHA-BCNN) [1] and deep associative memory [2]. Performance analysis is made with four

$$Acc = \sum_{i=1}^n \frac{CP_i}{TP_i} * 100 \quad (8)$$

From the above equation (8), accuracy of sentimental analysis for healthcare domain 'Acc' is measured based on the total predictions made for simulation purpose 'TP<sub>i</sub>'

distinct parameters namely, precision, recall, accuracy and error rate for respective number of tweets and distinct tweet sizes.

### 5.1. Scenario 1: Accuracy

In this section, the accuracy analysis is made with the aid of corona virus tweets and diabetes dataset. Accuracy is calculated as follows,

and the correct predictions made 'CP<sub>i</sub>' using the method for analysis. It is measured in terms of percentage.

2350

**Table 4: Average value of Accuracy using Diabetes and Corona Virus Tweets Dataset**

Datasets	Accuracy (%)		
	MSG-EDSA	MHA-BCNN	Deep Associative Memory
Diabetes dataset	86	81	79
Corona Virus dataset	79	76	73

With this the accuracy using MSG-EDSA method for diabetes dataset was said to be improved by 86 %, MHA-BCNN by 81% and deep associative memory by 79%. In MSG-EDSA method for corona virus tweets dataset the accuracy is improved by 79%, MHA-BCNN by 76% and deep associative memory by 73 %.

$$Pre = \frac{TP}{TP+FP} * 100 \quad (9)$$

### 5.2. Scenario 2: Precision

The second parameter of consideration taken into account is the precision. With all this factors into consideration, the precision is measured as given below.



From the above equation (9), precision '*Pre*' is estimated based on the true positive rate (i.e., correct analysis made using tweets) '*TP*'

and the false positive rate (i.e., falsely analysis the results using tweets) '*FP*'.

**Table 5: Average value of Precision using Diabetes and Corona Virus Tweets Dataset**

Datasets	Precision (%)		
	MSG-EDSA	MHA-BCNN	Deep Associative Memory
Diabetes dataset	88	83	79
Corona Virus dataset	77	74	70

It is worth noting that in Table 5, our method (MSG-EDSA) using diabetes and corona virus dataset show the best precision result compared to other state-of-the-art methods for the same samples collected from distinct users. This is specifically owing to the fact that our objective remains in maximizing the sentimental analysis factor and polarities in terms of the precision score, which combines the polarity accuracy on both positivity and negativity of the tweets. As polarity concentrates on large portion of the dataset, precision reflect the genuine performance of the classifier for sentimental analysis concerning healthcare domain. With this the

precision rate using MSG-EDSA for diabetes dataset was found to be improved by 88% MHA-BCNN by 83% and deep associative memory by 79%. In a similar manner, the precision rate using MSG-EDSA for corona virus tweets dataset was said to be improved by 77% and MHA-BCNN by 74% and deep associative memory by 70%.

2351

**5.3. Scenario 3: Recall**

In this section recall as a measure for sentimental analysis in the healthcare domain is formulated. This is measured as given below.

$$Rec = \frac{TP}{TP+FN} * 100 \tag{10}$$

From the above equation (10), recall rate '*Rec*' is measured based on the true positive '*TP*' and the false negative (i.e., falsely analyzing the tweets) '*FN*'.

**Table 6 : Average value of Recall using Diabetes and Corona Virus Tweets Dataset**

Datasets	Recall(%)



	MSG-EDSA	MHA-BCNN	Deep Associative Memory
Diabetes dataset	84	80	76
Corona Virus dataset	78	74	70

Table 6 reports the result of recall rate using MSG-EDSA, MHA-BCNN [1] and deep associative memory [2]. The initial apparent result is that deep learning method outperforms the state-of-the-art methods which verify the power of recurrent neural networks to capture direct association between tweets and classes. Comparing three different methods using two distinct datasets based on recall rate, our proposed MSG-EDSA method is the best among all. As the proposed MSG-EDSA method pays more focus to feature subset for selecting robust and relevant features, having higher performance selects the relevant feature for sentimental analysis with maximum recall rate. Moreover, the softmax classifier classifies the tweets and

produces the labels in a precise manner, therefore improving the recall rate using MSG-EDSA by 84%,MHA-BCNN by 80% and deep associative memory by 76% for diabetes dataset and MSG-EDSA by 78%,MHA-BCNN by 74% and deep associative memory by 70%, for corona virus tweets dataset.

2352

#### 5.4 .Scenario 4: Error rate

While predicting sentimental analysis for healthcare domain, a significant portion of error is said to occur by wrongly predicting the sentiments not being extremely positive as positive or extremely negative as negative and so on. The error rate in our work is mathematically expressed as given below.

$$Err = \frac{T_{WP}}{T_i} * 100 \tag{11}$$

From the above equation (11), the error rate 'Err' is measured based on the number of tweets wrongly predicted (i.e., extremely positive cases predicted with positive cases,

extremely negative cases predicted with negative and so on) 'T<sub>WP</sub>' to the tweet samples considered for simulations. It is measured in terms of percentage (%).

**Table 7: Average value of Error rate using Diabetes and Corona Virus Tweets Dataset**

Datasets	Error rate(%)		
	MSG-EDSA	MHA-BCNN	Deep Associative Memory
Diabetes dataset	13	15	17



Corona Virus dataset	14	16	18
----------------------	----	----	----

Finally, Table 7 given above shows the error rate evaluation by means of three different methods using diabetes and corona virus tweets dataset. Followed by which the actual classifications were made in Deep Recurrent Network, therefore reducing the error rate using MSG-EDSA method for diabetes dataset

by 13 % ,MHA-BCNN by 15% and deep associative memory by 17 %. In a similar manner, the error rate using MSG-EDSA method for corona virus tweets dataset reduced by 14%, MHA-BCNN by 16% and deep associative memory by 18 %.

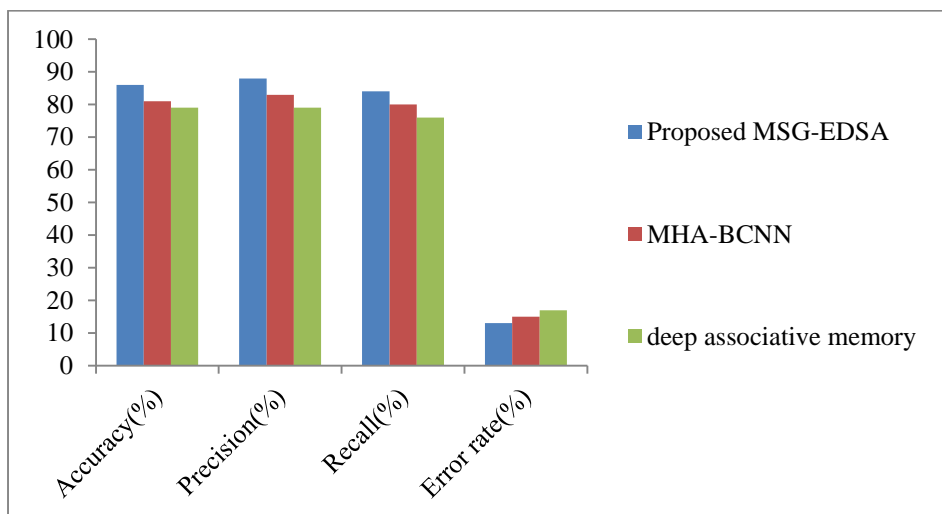


Figure 4: Graphical representation of Accuracy, Precision, Recall, Error rate using Diabetes dataset

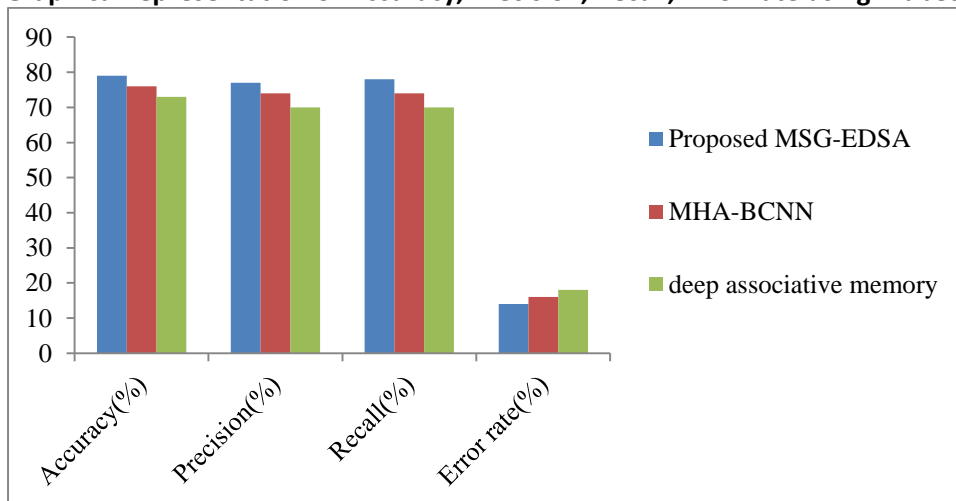


Figure 5: Graphical representation of Accuracy, Precision, Recall, Error rate using Corona Virus Tweet dataset



Figure 4 and 5 given above shows the performance analysis of accuracy, precision, recall, error rate using diabetes dataset and corona virus tweets dataset. The proposed MSG-EDSA method is compared with existing two methods such as MSG-EDSA and MHA-BCNN methods. Compared to all methods proposed technique achieve better performance in terms of accuracy, precision, recall, error rate.

## 6. CONCLUSION

Sentiment analysis in healthcare domain permits healthcare providers to classify patients' comments based on their tweets from the Twitter social media. On the basis of the scoring comments on whether they are extremely positive or positive, extremely negative or negative, neutral can quantify patients' experience and identify areas of improvement. In this work, an effective sentimental analysis method using Mean Silhouette-based Genetic and Elman Deep Sentiment Analysis (MSG-EDSA) is proposed. The different processes involved in the design are pre-processing, feature selection and robust classification. First, raw user samples including distinct attributes are pre-processed. Next, with the pre-processed attributes and tweets, precise and relevant features are selected using the Mean Silhouette-based Genetic Feature Selection algorithm. Finally, the selected features are provided as input to the Elman Deep Recurrent Network Sentiment Classification algorithm. The proposed MSG-EDSA method is simulated using Python with the aid of diabetes and corona virus tweets dataset. The simulation consequences validated that the MSG-EDSA method provides better results in performance metrics like, accuracy, precision, recall and error rate compared to the state-of-the-art methods.

## REFERENCES

[1] Kodati Dheeraj, TeneRamakrishnudu, "Negative emotions detection on online mental-health related patients texts using the

deep learning with MHA-BCNN model ", Expert Systems With Applications, Elsevier, May 2021[Multi Head Attention with Bidirectional long short-term memory and Convolutional Neural Network (MHA-BCNN)

[2] HananeGrissette, El Habib Nfaoui, "Deep associative learning approach for bio-medical sentiment analysis utilizing unsupervised representation from large-scale patients' narratives", Personal and Ubiquitous Computing, Springer, Jun 2021 [deep associative memory]

[3] Sharon Chekijian, Huan Li, SamahFodeh, "Emergency care and the patient experience: Using sentiment analysisand topic modeling to understand the impact of the COVID-19pandemic", Health and Technology, Springer, Aug 2021

[4] Cristóbal Colón-Ruiz, Isabel Segura-Bedmar, "Comparing deep learning architectures for sentiment analysis on drugreviews", Journal of Biomedical Informatics, Elsevier, Aug 2020

[5] Elia Gabarron, Enrique Dorronzoro, Octavio Rivera-Romero, Rolf Wynn, MD, "Diabetes on Twitter:A Sentiment Analysis", Journal of Diabetes Science and Technology, Jul 2019

[6] DemijanGrgic, VedranPodobnik, "Application of Data Science for UnderstandingEmotional Dimensional Behavior and TheirConnection to Uncertainty and Risk Behavior", IEEE Access, May 2021

[7] Nikos Fazakis, Otilia Kocsis, Elias Dritsas, Sotiris Alexiou, Nikos Fakotakis, Konstantinos Moustakas, "Machine Learning Tools for Long-Term Type 2 Diabetes Risk Prediction", IEEE Access, Jul 2021

[8] Nalini Chintalapudi, Gopi Battineni,Marzio Di Canio, GetuGamoSagaro, Francesco Amenta, " Text mining with sentiment analysis on seafarers' medical documents", International Journal of Information Management Data Insights", Elsevier, Jul 2021

[9] María del Pilar Salas-Zárate, JoséMedina-Moreira, Katty Lagos-Ortiz,Harry Luna-Aveiga,

2354



Miguel Ángel Rodríguez-García, Rafael Valencia-García, “Sentiment Analysis on Tweets about Diabetes: An Aspect-Level Approach”, Computational and Mathematical Methods in Medicine, Hindawi, Feb 2017

[10] Quynh C. Nguyen, Kimberly D. Brunisholz, Weijun Yu, Matt McCullough, Heidi A. Hanson, Michelle L. Litchman, Feifei Li, Yuan Wan, James A. VanDerslice, Ming Wen, Ken R. Smith, “Twitter-derived neighborhood characteristics associated with obesity and diabetes”, Scientific Report, Nov 2017

[11] Edmund Kwong Wei Leow, Binh P. Nguyen, Matthew Chin Heng Chua, “Robo-advisor using genetic algorithm and BERT sentiments from tweets for hybrid portfolio optimization”, Expert Systems With Applications, Elsevier, Apr 2021

[12] Ruchi Mittal, Amit Mittal, Ishan Aggarwal, “Identification of affective valence of Twitter generated sentiments during the COVID-19 outbreak”, Social Network Analysis and Mining, Springer, Jul 2021

[13] Farrukh Aslam Khan, Khan Zeb, Mabrook Al-Rakhami, Abdelouahid Derhab, Syed Ahmad Chan Bukhari, “Detection and Prediction of Diabetes Using Data Mining: A Comprehensive Review”, IEEE Access, Mar 2021

[14] Beatrice Alex, Donald Whyte, Daniel Duma, Roma English Owen and Elizabeth A. L. Fairley, “Classifying patient and professional voice in social media health posts”, BMC Medical Informatics & Decision Making, Sep 2021

[15] Mrityunjay Singh, Amit Kumar Jakhar, Shivam Pandey, “Sentiment analysis on the impact of coronavirus in social life using the BERT model”, Social Network Analysis and Mining, Springer, Feb 2021

[16] Qi Yao, Rita Yi Man Li, Lingxi Song, M. James C. Crabbe, “Construction safety knowledge sharing on Twitter: A social network analysis”, Safety Science, Elsevier, Jul 2021

[17] Tomoo Hidaka, Noriko Takahashi, Katsue Hashimoto, Mariko Inoue, Yukiko Terada, Shota Endo, Takeyasu Kakamu, Taro Tsukahara, Koichi Abe, Tetsuhito Fukushima, “Qualitative and Quantitative Study on Components of Future Time Perspective and Their Association with Persistent Treatment for Type 2 Diabetes”, Diabetes Therapy, Springer, Oct 2021

[18] Atae Rezaei Aghdam, Jason Watson, Shah Jahan Miah, Cynthia Cliff, “A novel information sharing framework for people living with type-2 diabetes in the context of a group education program”, Health Information Science and Systems, Springer, Aug 2021

[19] Hanaa Torkey, Elhossiny Ibrahim, EZZ El-Din Hemdan, Ayman El-Sayed, Marwa A. Shouman, “Diabetes classification application with efficient missing and outliers data handling algorithms”, Complex & Intelligent Systems, Springer, Mar 2021

[20] Fatima-Ezzahra Lagrari, Youssfi Elkettani, “Traditional and Deep Learning Approaches for Sentiment Analysis: A Survey”, Advances in Science, Technology and Engineering Systems Journal, Sep 2021

2355

