[1]Yongfang Zhang
[2*]Hongxing Fan

Application of Multimodal Data Fusion Attentive Dual Residual Generative Adversarial Network in Sentiment Recognition and Sentiment Analysis

Journal of Electrical Systems

Abstract: - Recent advancements in Internet technology have led to increased multi-modal data posting on social media, online shopping portals, and video repositories recognizing significance of inter-modal utterances before combining multiple modes. In this manuscript, Application of Multimodal Data Fusion Attentive Dual Residual Generative Adversarial Network in Sentiment Recognition and Sentiment Analysis (MDF-DRGAN-SR-SA) is proposed. The input data are collected from CMU-MOSI dataset. Initially the input data is preprocessed using Subaperture Keystone Transform Matched Filtering (SAKTMF) is used to clean unwanted data. Then, feature extraction is done by Two-Sided Offset Quaternion Linear Canonical Transform (TSOQLCT) to extract unimodal features likes acoustic, textual, visual. Then the selected features are given to ADRGAN classifying Sentiment Recognition and Sentiment Analysis likes positive, negative, neutral. In general, ADRGAN doesn't express some adaption of optimization strategies for determining optimal parameters to assure accurate classification of Sentiment Recognition and Sentiment Analysis. Hence, Northern Goshawk Optimization Algorithm (GOA) is proposed to enhance weight parameter of ADRGAN, which precisely classifies the Sentiment Recognition and Sentiment Analysis in positive, negative and neutral. The proposed model is implemented and its efficiency is evaluated utilizing some performance metrics likes accuracy, precision, specificity, sensitivity,F1-score. The MDF-DRGAN-SR-SA method provides 25.85%, 26.79% and 27.63% higher accuracy; 35.66%, 34.97% and 26.57% higher precision; 28.18%, 29.52% and 25.68% higher specificity is compared with existing method such as Two-Level Multimodal Fusion for SA in Public Security (TMDF-SA-PS), Multimodal SA Depend on Adaptive Modality-Specific Weight Fusion Network (MFN-SA-AMW) and Multimodal SA Utilizing Multi-tensor Fusion Network and Cross-modal Modeling(MTFN-SA) respectively.

Keywords: Attentive Dual Residual Generative Adversarial Network, Northern Goshawk Optimization Algorithm, Subaperture Keystone Transform Matched Filtering, TSOQLCT.

I. INTRODUCTION

The social media sites, online shopping, travel conductor websites, video repositories have all grown recently, and this has allowed users to share their opinions about entities and products in textual, audiovisual, or consolidation of two/more modalities at once [1].Several businesses, online pornographers use data to assess customer complaints, learn about people's thoughts and feelings regarding their products or services, and more [2, 3]. Customers' quality of life is enhanced by the information gleaned from audio-visual content, which enables them to make better decisions about what to buy, which services to use, what movies to watch, where to go on vacation, and other matters. Textual modality was the only place where classical sentiment analysis and emotion detection could be applied [4]. In contrast, researchers have recently been quite interested in multimodal sentiment analysis, affective computing [5]. Main benefit of evaluating multimodal content over unimodal content that allows for the use of a wider range of information from many modalities, which develops classification accurateness of system [6, 7]. When analyzed with a system that solely used textual modalities, accessibility of facial expressions, vocal cues strengthens the system and increases its accuracy in e-tourism, diction, e-health [8-10]. The results show that the proposed bimodal emotion recognition system achieves higher accuracy than unimodal-depends schemes when the audio-visual characteristics are concatenated to a bimodal feature vector [11]. Subsequent researchers fused features taken from distinct modalities to create bimodal and tri-modal audio-visual sentiment analysis and emotion recognition models using either early feature fusion/late fusion or hybrid fusion.The 3D activation valance approach is used to extract verbal and auditory characteristics for the purpose of classifying emotions [12-15]. Textual-visual-acoustic features are retrieved using convolutional neural networks, multi-kernel learning is employed to analyze sentiment and emotion as classifier [16]. Sentiment analysis, emotion categorization have seen introduction of memory fusion and tensor fusion algorithms more recently [17].

In order to take into account the significance of modality prior to attention-based inter-modality fusion and utterance-level contextual information, the suggested strategy takes into account the influence of nearby utterances [18, 19]. In multimodal sentiment analysis, emotion identification, videos are typically segmented into brief opinion chunks, or utterances. Every speech is carefully assigned an emotion or sentiment label. Each utterance is examined separately in utterance-level sentiment analysis, regardless of its nearby utterances [20]. Since a movie consists of a series of opinion segments, the present utterance's outcome is influenced by nearby

[1]College of Educational Science, Yan'an University, Shaanxi, Yanan, China, 716000
[1]E-mail: yongfang2025@163.com
[2]School of Technology, Fuzhou Technology and Business University, Fuzhou, Fujian, 350715, China
*Corresponding Author Email: 18060514524@163.com

utterances (contextual information). Conventional fusion techniques and utterance-level sentiment analysis are unable to extract context from several utterances [21].

A. Problem statement and Motivation

Context cannot be extracted from multiple utterances using sentiment analysis or traditional fusion techniques. Multimodal sentiment analysis with context modelling focuses on contextual feature vector. Such method focuses on modality-exact contextual features, limiting influence of other modalities. To overcome these problems, deep learning algorithm is used. These are inspired to do this research.

A Multimodal Data Fusion with Attentive Dual Residual Generative Adversarial Network can better predict sentimental analysis by combining multi-modal features. Initially, acoustic, textual, visual features are extracted from TSOQLCT. Cross-modal modelling is utilized to extract context of multimodal features. Attain fine-grained bimodal relationship features, three bimodal cross-attention values are a selection of each pair of modalities. The Multimodal Data fusion ADRGAN with cross-modal modelling captures dynamic information from intra- and inter-modal interactions, resulting in more accurate sentimental prediction.

B. Contribution

Major contribution of this investigation work is brief below

- To handle the characteristics from many modalities and nearby utterances in multimodal sentiment analysis, emotional computing, an efficient attentive method is suggested.
- The model shows that contextual utterances inside the modality and attentive inter-modality features can be combined to produce superior multimodal feature representation.
- The most advanced model for classifying sentimental analysis is demonstrated and evaluated using CMU-MOSI dataset.

Continuing paper is arranged as below: unit 2 presents the literature review, unit 3describes the proposed method, unit 4 shows result and discussions, unit5 gives the conclusion.

II. LITERATURE REVIEW

Among the various works on deep learning depend sentimental analysis; some of the latest investigations are revised here,

Sun et al. [22] have presented TMDF for SA in PS. In the presented method process Multimodal data was merged using fusion processes to ensure accuracy of the data utilized for after-classification tasks. To complete the sentiment analysis task, suggests two-level multimodal fusion technique that combines decision-level, data-level fusion. The presented method was tested on CMU-MOSI, CMU-MOSEI, IEMOCAP dataset, with experimental findings, studies on ablation confirming TlMF's efficiency in recording useful data from all test methods. It provides high precision and it provides low specificity.

Zhang et al. [23] have presented Multimodal SA Depend on AMW Fusion Network. In the presented method suggests adaptive modality-specific weight fusion network to solve problems encountered during multimodal information fusion. It employs different weight calculation techniques at numerous points of method. Such method's testing, verification phase, learned weight-mapping system was employed to calculate weights of numerous methods. The multiple modal vectors of features were analyzed with enhanced to ensure that outcomes of fusion preserve the modality's distinctiveness while getting multiple modal data interaction. AdaMoW was validated on CMU-MOSI, CMU-MOSEI datasets. It provides high sensitivity and it provides low F1-score.

Yan et al. [24] have presented Multimodal SA Utilizing MTFN by Cross-modal Modeling. In the presented method suggests a little swift growth of social media platforms, an increasing number of people are using videos on the internet to express their emotions and opinions. Multimodal communication extraction of features by cross-modal was used to identify connection of emotional data across various modes. Furthermore, multi-tensor fusion network was utilized to simulate relationship between several bidirectional connects and to forecast emotional state of multifaceted features. It executes well in both classification, regression tests on dual public dataset CMUMOSI, CMU-MOSEI. It provides high accuracy and it provides low sensitivity.

Al-Tameemi et al. [25] have presented Multi-Model Fusion Framework Utilizing DL for Visual-Textual Sentiment Classification. Here, suggested framework consists of three DNNs. Two distinct neural networks have been suggested for extracting most emotionally significant features of textual, visual data. As a consequence, further discriminating features were collected to ensure accurate sentiment classification. A comprehensible visual-textual sentiment classification framework was refined by utilizing Local

Comprehensible System-agnostic Explaining method to guarantee method's clearness, robustness. It provides high precision and it provides low accuracy.

Younis et al. [26] have presented Evaluating Ensemble Learning Techniques for Multi-Modal Emotion Realization Utilizing Sensor Data Fusion. The presented method combines on-body emotion measurements, sensory data, as well as to achieve these objectives: Gathering a multifaceted data set that includes environmental, physiological, and emotional responses. To create accurate predictive models, data was gathered from participants walking around the Minia University campus. It provides high specificity and it provides low sensitivity.

Njok et al. [27] have presented DL Depend Data Fusion Techniques for Multimodal Emotion Recognition. The suggested Data fusion was important stage in multimodal emotion recognition due to precision of recognition models was heavily dependent on how the various modalities are paired. Examine the efficacy of neural network (DL) models in combining data and multifaceted recognition of emotions. The benefits of this paper are two fold DL for multiple modal integration and categorization: early fusion, hybrid fusion, multi-task learning. It provides higher accuracy, lower precision.

Han et al. [28] have presented Refining Multimodal Fusion and Hierarchical Mutual Information Maximization for Multimodal Sentiment Analysis. Here, suggested Multi Modal InfoMax (MMIM) was a structure that hierarchically maximizes Mutual Information within unimodal input pairs along with among multimodal combination results, unimodal input to maintain task-linked data via multifaceted fusion. The framework was trained alongside main task to develop downstream MSA task's performance. To tackle difficult problem of MI limits, create a group of computationally simple parametric, non-parametric techniques for estimating truth value. It attains higher precision, lower specificity.

II. PROPOSED METHODOLOGY

In this section, Application of Multimodal Data Fusion Attentive Dual Residual Generative Adversarial Network in Sentiment Recognition and Sentiment Analysis (MDF-DRGAN-SR-SA) is proposed. The block diagram of proposed MDF-DRGAN-SR-SA approach is represented in Figure 1. This process involves five steps likes data acquisition, pre-processing, feature extraction, classification, optimization. The detailed description of all steps given below,

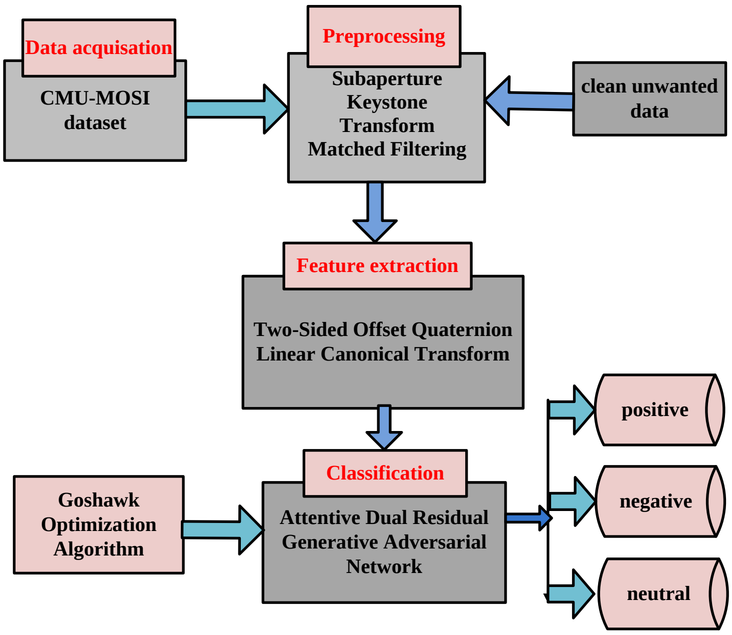

Figure1: Block Diagram for proposed MDF-DRGAN-SR-SA method

A. Data acquisition

The data is gathered through CMU-MOSI dataset [29] and videos from YouTube. The dataset comprises ninety-three videos with eighty-nine different voices evaluating various products. All video reviews contain several opinion sections known as utterances. The CMU-MOSI dataset contains 2199 opinion segments, or utterances. Each segment or utterance lasts an average of 4.2 seconds and contains 12 words. Five assessors manually labelled each utterance at -3 to +3. -3 and +3 ranges denotes intensely negative, intensely positive polarities. Final label for an utterance is based on the average vote. In accordance with current multimodal sentiment analysis research, positive and negative polarity only used. Then, collected data are fed to pre-processing.

B. Pre-processing using Subaperture Keystone Transform Matched Filtering

In this section, the acquired data are pre-processed using SAKTMF [30] technique is used to clean the unwanted data. SAKTMF are methods for pre-processing improving data quality. The SAKTMF divides the aperture into sub apertures, and matched filtering is a technique for improving data quality by correlating received data with known patterns. In data processing, normalization denotes to process of adjusting range of values in dataset to a standard or common scale. It aids in making disparate datasets comparable and alleviates issues caused by different scales or units. Then, resizing data are calculated in the equation (1).

$$Dd(e_s, N_w) = \xi resi[\frac{e_s}{dI}]resi[\frac{e_i}{e_s+e_i}\frac{N_w}{N_h}]\ Dd \tag{1}$$

where D is the data analysis, e_s is changing the size of data, N_w is changing the width of data, $\xi resi$ is the resizing of data, $\frac{e_s}{dI}$ is changing the size of data in sentimental analysis, e_i is changing the data size, $\frac{e_i}{e_s+e_i}$ is the calculation of changes in shape and size of data, $\frac{N_w}{N_h}$ is the calculation of changes in data. Then the sentimental analysis of data is calculated in equation (2).

$$Dd(e_s, N_w) = \sum_{a=1}^{s} Dd_a(Dd(e_s, N) \tag{2}$$

where D is the data analysis, e_s is changing the size of data, N_w is changing the width of data, $\sum_{a=1}^{s}$ is the calculation of sentimental analysis, Dd_a denotes analysis of data, N signifies number of data, $Dd(e_s, N)$ denotes number of analysis data. Smoothing Edges detection and removal of noise is calculated in equation (3).

$$Dd_g(r_n, N_{w,s}) = Dd_s(e_{g\ s}, N_{w,n}) \otimes \Delta\varphi(e_s, N_{w,s}) \tag{3}$$

where Dd_g is the Smoothing Edges analysis data, r_n is the removal of data, Δ is the processed data, $\otimes$ is the data calculation, φ is the sigmoid calculation of value. Then clean the unwanted data are calculated in equation (4).

$$Dd_{s,seg}(e_s, N_{w,s}) = Dd_i(e_s, N_{w,s}) resi[-i\frac{3\pi}{\gamma} y_n It_d] \tag{4}$$

where $Dd_{s,seg}$ the clean unwanted data, e_s is changing the size of data, $N_{w,s}$ is width and size of data, Dd_i is the analysis of data identification, $resi$ is the resize of data, $\frac{3\pi}{\gamma}$ is the matrix of value calculation, y_n is the remove unwanted data, It_d is the time taken of clean data. Finally, SAKTMF method cleaned the unwanted data. Then pre-processed data is given into feature extraction.

C. Feature extraction using two-sided offset quaternion linear canonical transform

TSOQLCT [31] is presented for feature extraction. TSOQLCT to enhance detecting the image objects. It's a method for improving feature extraction and sentimental analysis in data. Methods of feature extraction seek to identify and extract relevant information or patterns from data that are indicative of specific conditions or characteristics. Feature extraction techniques may be used in the context to identify specific sentimental requirements and analysis. To increase the accuracy in the original, a new technique for the feature extraction process is presented TSOQLCT. Then the sentimental analysis is calculated in equation (5).

$$E_P\{e\}(v) = \frac{1}{\sqrt{(2r)^2}} \int f^{-cs_1v_i} e(s) f^{-ds_2v_2} \tag{5}$$

where E_P are the sentimental recognition, $\{e\}$ is the changes of data, (v) is the variations of analysis, $\frac{1}{\sqrt{(2r)^2}}$ is the calculation of ratio of sentimental analysis, $f^{-cs_1v_i}$ is the identification of sentimental recognition variation, $f^{-ds_2v_2}$ is the sentimental recognition value variation. Then the number of sentimental recognition is calculated in equation (6).

$$\Delta n_h^2 = (\oint n_h^2 |e(n)|^2 dn \quad (6)$$

where Δn_h^2 the number of sentimental recognition, $\oint n_h^2$ is the total calculation of sentimental analysis, $e(n)$ is the sentimental changes in data, dn is the data quality analysis. The proposed TSOQLCT is used to extract unimodal features such as acoustic, textual, and visual. The Audio feature extraction, voice normalization, and intensity thresholding of the acoustic is given by the Equation (7)

$$a = \langle a_1, a_2, a_3, ..., a_n \rangle \quad (7)$$

where a denotes acoustic, a_1 signifies acoustic value one, a_2 is the acoustic value two, a_3 is the acoustic value three, a_n is the e number of acoustic analysis. The acoustic analysis is manually transcribed to obtain textual data is given by the Equation (8).

$$t = \langle t_1, t_2, t_3, ..., t_n \rangle \quad (8)$$

where t is the textual, t_1 is the textual value one, t_2 is the textual value two, t_3 is the textual value three, t_n is the e number of textual analysis. Similar process is utilized to extract visual features is given in equation (9).

$$v = \langle v_1, v_2, v_3, ..., v_n \rangle \quad (9)$$

where v is the visual, v_1 is the visual value one, v_2 is the visual value two, v_3 is the visual value three, v_n is the number of visual analysis. Finally, TSOQLCT is extracted the features such as acoustic, textual, and visual. After completing feature extraction, extracted features are given to ADRGAN.

D. Sentimental analysis Classification utilizing Attentive Dual Residual Generative Adversarial Network

ADRGAN [32] is discussed. ADRGAN is used for classifying Sentiment Analysis such as positive, negative and neutral. This model uses ADRGAN as base classifiers to efficiently analyze the Multimodal Data Fusion. The data analytics lifecycle for Multimodal Data Fusion involves aggregating individual classifier predictions to reduce variance. In the ADRGAN of Sentiment Analysis this could entail combining data from multiple sensors to develop accuracy of Multimodal Data Fusion of sentimental analysis. ADRGAN is an application that uses advanced machine learning techniques to develop performance of Multimodal Data Fusion. Then quality of sentimental analysis is calculated in equation (10).

$$U(Q, A) = P_{L-R_s}[\log(A(L))] + P_{J-R_s-s}[\log(1 - A(Q(j))] \quad (10)$$

where U is the is the Data Fusion, $U(Q, A)$ is the quality and analysis of the sentimental, P is the parameter value, P_{L-R_s} is the parameter value of sentimental analysis, $A(L)$ is the sentimental analysis variation, j is the input value, P_{J-R_s-s} is the calculation of parameter in sentimental analysis, $A(Q(j)$ is the quality and analysis. Then, sentimental analysis of the positive is calculated in equation (11).

$$U_r = L_{D-P}(Q(J), S) - (Q(J) * W \quad (11)$$

where U_r denotes analysis of sentimental is positive, L_{D-P} is the calculation to the parameter analysis, $Q(J)$ is the quality of the sentimental analysis, S is the analysis in the recognition, $(Q(J) * W$ is the quality of the sentimental analysis in positive. Then, sentimental analysis of the negative is calculated in equation (12).

$$P(sm) = \sum_{j=0}^{w} \left| x^{(j)} \right| - \left| y^{(j)} \right| \quad (12)$$

where $P(sm)$ the parameter value of sentimental in negative, $\sum_{j=0}^{w}$ is the calculation of negative analysis, $x^{(j)}$ is analysis of the input value, $y^{(j)}$ is analysis of the input character. Then, sentimental analysis of the neutral is calculated in equation (13).

$$Q^* = U(Q,D) + (s_1.P_L + s_2.L_P) \tag{13}$$

where Q^* the quality analysis in neutral, U is the Data Fusion, D is the data, $U(Q,D)$ is the Data Fusion analysis accuracy, $s_1.P_L$ represent the parameter value of first sentimental analysis, $s_2.L_P$ represent the parameter value of first sentimental analysis. Finally, ADRGAN is used for classified the Sentiment Analysis such as positive, negative and neutral. In this work, Northern Goshawk Optimization (NGO) is employed to enhance ADRGAN. Here, NGO is applied for tuning weight, bias parameter of ADRGAN.

E. Optimization using Northern Goshawk Optimization

In this section, NGO [33] process is used to optimize the weight parameter of ADRGAN. NGO used to optimize ADRGAN weight parameters which effectively analyse the sentimental such as positive, negative and overloading. The optimized weight parameters obtained through the NGO driven optimization process are then applied within the ADRGAN model to improve sentimental Recognition and analysis classification performance. The NGO algorithm principle can be divided into five major stages, which are includes in following stages.

Step 1: Initialization

Initialization weight parameter is using the model is expressed in the form of sentimental analysis which is used to calculate the modulation detection in weight parameters of ADRGAN is calculated in the equation (14).

$$s = \begin{bmatrix} s_{1,1} & s_{1,2} & \cdots & s_{1,s} \\ s_{2,1} & s_{2,2} & \cdots & s_{2,s} \\ \vdots & \vdots & \vdots & \vdots \\ s_{s,1} & s_{s,2} & \cdots & s_{a,s} \end{bmatrix} \tag{14}$$

where s the parameter weight calculation in the sentimental, $s_{a,s}$ is the number of parameter weight calculation in the sentimental analysis.

Step 2: Random generation

After initialization process, U_r, Q^* weight parameter is created at random through ADRGAN method.

Step 3: Fitness Function

The fitness function for optimizing weight parameters in sentimental analysis using machine learning typically involves measuring the model Multimodal Data Fusion align with the actual outcomes truth related to analysis. The fitness function measures the performance of the model and guides the optimization process in adjusting the weight parameters. The Fitness function is calculated using eqn (15).

$$FitnessFunction = [optimizing\ U_r, Q^*] \tag{15}$$

Step 4: Exploration phase

Exploration phase is used to evaluate the weight parameter in optimization. The ADRGAN weight parameter can be used to improve the calculation. Then, calculates the evaluation of each optimization vector is given in the equation (16).

$$y_{(i,j)}^{L1} = y_{i,j} + D \times (U \times L - 1) y_{i,j} + L \tag{16}$$

where $y_{(i,j)}^{L1}$ represent analyze the parameter weight, $y_{i,j}$ is the sentimental analysis, D is the data, U is the Data Fusion, $L-1$ is the sentimental weight parameter calculation.

Step 5: Exploitation phase for optimizing U_r, Q^*

The term exploitation of weight parameter calculation denotes to process of optimizing method's weights to improve its accuracy in detecting Multimodal Data Fusion using machine learning models proposed method is mathematically expressed equation (17).

$$y_{i,j} = \{ \begin{matrix} y_i^{L1}, D_i^s < D_i \\ y_j^{P1}, D_j^s < D_j \end{matrix} \tag{17}$$

where y_i^{L1} is the sentimental data analysis, D_i^s is the Multimodal Data Fusion analysis, y_j^{P1} is the parameter analysis, D_j^s is the Multimodal Data Fusion recognition. Figure 2 shows that the flowchart of NGO for optimizing ADRGAN parameter

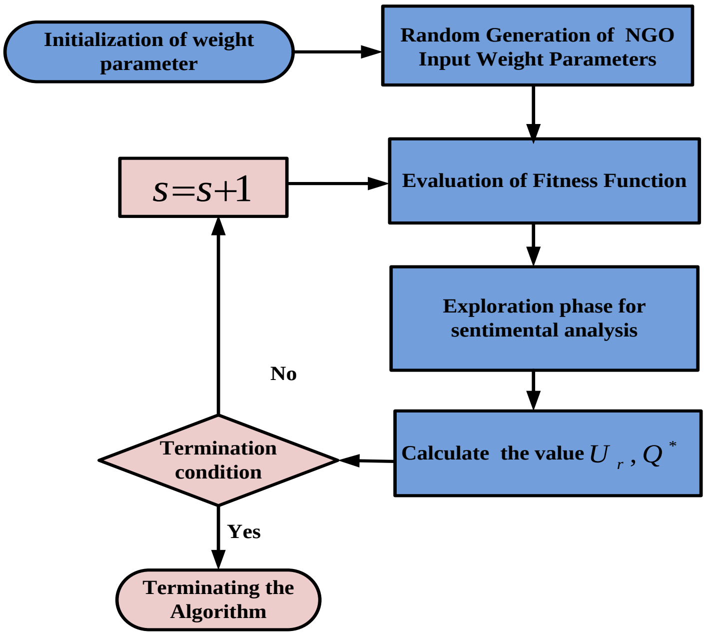

Figure 2: Flowchart of NGO for optimizing ADRGAN parameter

Step 6: Termination

Finally, the factor U_r, Q^* is optimized by NGO; will repeat step 3 till it reaches halting criteria $s = s + 1$. ADRGAN is optimized with NGO effectively for detection review with better accuracy. Thus the proposed MDF-DRGAN-SR-SA method effectively analyzes Tennis Teaching with higher accuracy and lower Mean Squared Error.

IV. RESULT WITH DISCUSSION

The simulation outcomes of MDF-DRGAN-SR-SA are discussed. The proposed approach is activated in Python utilizing mentioned metrics. The performance of the MDF-DRGAN-SR-SA technique is evaluated with existing techniques, likes TMDF-SA-PS,MFN-SA-AMW, and MTFN-SA respectively.

A. Performance Metrics

It is evaluated to validate the efficiency of MDF-DRGAN-SR-SA technique. Following confusion matrix is necessary for that.

- TP: Instances are actually positive then categorized sentimental analysis as positive.
- TN: Instances is actually negative then categorized sentimental analysis as negative.
- FP: Instances is actually negative then categorized sentimental analysis as positive.
- FN: Instances is actually negative then categorized as sentimental analysis as neutral.

1) Accuracy

Accuracy is used to evaluate a detection system's presentation in correctly identifying and classifying instances within a given dataset, categorized as sentimental recognition and analysis and is calculated by eqn (18),

$$Accuracy = \frac{(TP + TN)}{(TP + FP + TN + FN)} \tag{18}$$

2) Precision

Precision measures the system's ability to detect positive cases correctly out of all predicted positive cases. It is proportion of true positives to total false positives, true positives. It is determined by eqn (19),

$$\Pr ecision = \frac{TP}{(TP + FP)} \tag{19}$$

3) Sensitivity

The proportion of appropriately forecast positive instances among each actual positive instances is deliberate as sensitivity. It is assesses ability to identify positive instances correctly and is scaled by eqn (20),

$$sensitivity = \frac{TP}{TP + FN} \tag{20}$$

4) Specificity

It weighs the proficiency in single class by approximating the probability that is computed by eqn (21),

$$Specificity = \frac{(TN)}{(FP + TN)} \tag{21}$$

5) F1-score

It denotes harmonic mean of precision, sensitivity. It is widely utilized as evaluation metric in binary, multi-class classification, combines precision, sensitivity into single metric to gain best understanding of method performance. It is determined by eqn (22).

$$TPR = \frac{TP}{TP + FN} \tag{22}$$

B. Performance analysis

Figure 3 to 7 shows simulation result of accuracy, precision, specificity, sensitivity and F1-Scoreare analyzed for the proposed MDF-DRGAN-SR-SA technique is analyzed with existing technique likes TMDF-SA-PS, MFN-SA-AMW, and MTFN-SA respectively.

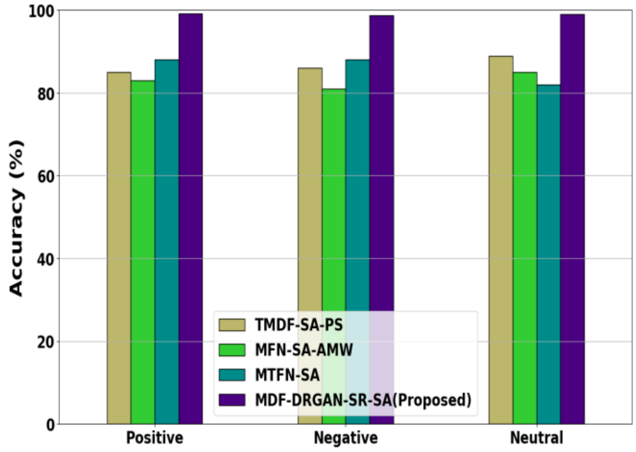

Figure 3: Accuracy analysis

Figure 3 depicts accuracy analysis. The MDF-DRGAN-SR-SA method sentimental analysis provides the 28.46%, 21.34 and 33.81% higher accuracy for positive, 22.55%, 24.72% and 29.63% higher accuracy for negative and 25.18%, 21.52% and 28.68% higher accuracy for neutral are analyzed with existing method likes TMDF-SA-PS, MFN-SA-AMW, and MTFN-SA respectively.

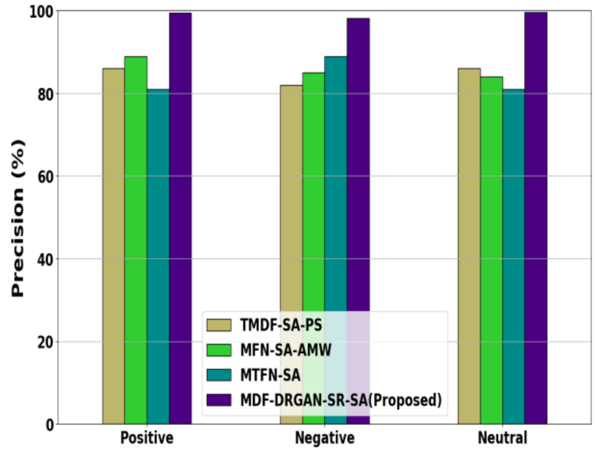

Figure 4: Precision analysis

Figure 4 depicts precision analysis. The MDF-DRGAN-SR-SA method sentimental analysis provides 22.88%, 26.52% and 34.63% higher Precision for positive, 32.66%, 34.97% and 29.57% higher Precision for negative and 29.12%, 29.33%, 32.55% greater precision for neutral are analyzed with existing method such as TMDF-SA-PS, MFN-SA-AMW, and MTFN-SA respectively.

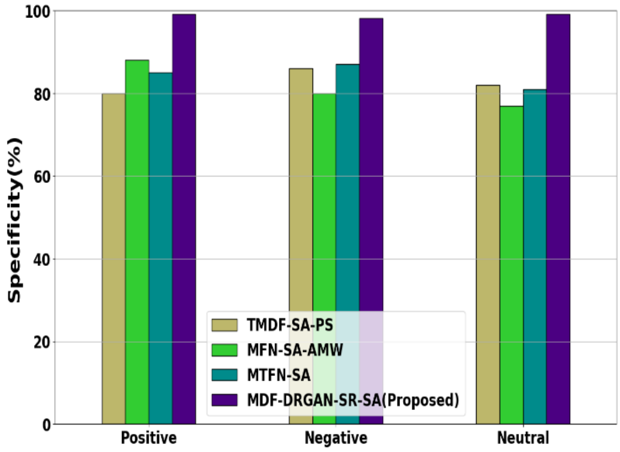

Figure 5: Specificity analysis

Figure 5 depicts specificity analysis. The MDF-DRGAN-SR-SA method sentimental analysis provides the 28.46%, 21.34 and 33.81% higher specificity for positive, 22.55%, 24.72% and 29.63% higher specificity for negative and 28.34%, 23.29% and 21.57% higher specificity for neutral are analyzed with existing method such as TMDF-SA-PS, MFN-SA-AMW, and MTFN-SA respectively.

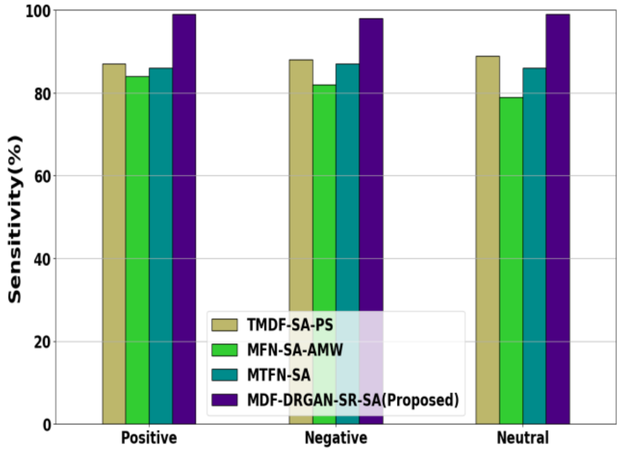

Figure 6: Sensitivity analysis

Figure 6 shows sensitivity analysis. The MDF-DRGAN-SR-SA method sentimental analysis provides achieves an improvement of 22.37%, 27.89%, and 31.37% higher sensitivity for positive, 23.47%, 28.76%, and 24.67% higher sensitivity for negative and 22.72%, 28.34% and 33.29% higher sensitivity for neutral are analyzed with existing technique likes TMDF-SA-PS, MFN-SA-AMW, and MTFN-SA respectively.

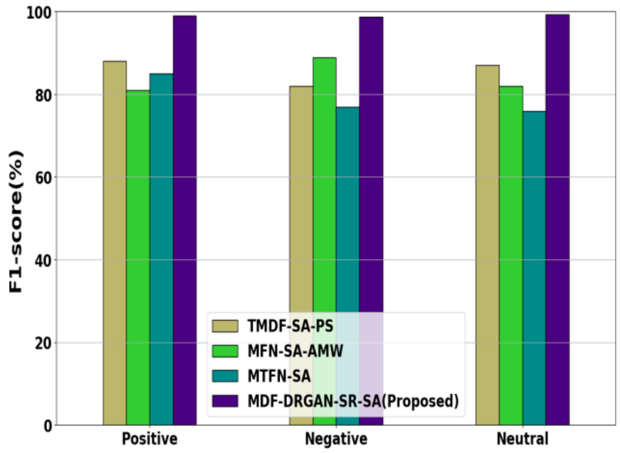

Figure 7:F1-score analysis

Figure 7 depicts F1-score analysis. The MDF-DRGAN-SR-SA method sentimental analysis provides achieves an improvement of 19.52%, 25.65%, and 29.82% higher F1-Score for positive, 26.54%, 19.95%, and 20.83% higher F1-Score for negative and33.29%,18.57% and 20.67%greaterF1-score for neutral are analyzed with existing method likes PDTI- MSDF-HDL, RDSA-MSDF and ODA-MAFF-CVE methods respectively.

C. Discussion

DRGAN model is trained and tested by cleaned the data approaches. During the experiment, changed the hyper parameters that shape the DRGAN-Change a single parameter and the others will be fixed. It also successfully developed DRGAN-models that classify sentimental analysis into positive, negative and neutral. The sentimental recognition and analysis accuracy of MDF-DRGAN-SR-SA is 28.46%, 21.34 and 33.81% higher than existing methods such as TMDF-SA-PS, MFN-SA-AMW, and MTFN-SA respectively. Similar to this, the precision of proposed method is 97.94% analyzed with sensitivity of comparison techniques of 87.56%. The proposed method MDF-DRGAN-SR-SA has high F1-score and specificity evaluation metrics than existing methods. Therefore, the comparative methods are expensive than the proposed technique. As a result, the proposed technique classifies sentimental recognition and analysis is effectively and efficiently.

IV. CONCLUSION

The proposed, Application of Multimodal Data Fusion Attentive Dual Residual Generative Adversarial Network in Sentiment Recognition and Sentiment Analysis (MDF-DRGAN-SR-SA) has successfully implemented. The proposed technique describes Contextual information extraction and multimodal fusion is important components of multimodal sentiment analysis, recognition. The MDF-DRGAN-SR-SA technique is activated in Python. The mentioned performance metrics is analyzed. The MDF-DRGAN-SR-SA method provides 29.86%, 38.89 and 35.99% higher sensitivity; 27.67%, 33.74%, and 34.56% greaterF1-score is analyzed with existing method likes TMDF-SA-PS, MFN-SA-AMW, MTFN-SA methods respectively.

Acknowledgement:

The 13th Five-Year Plan of National Education Science the Key Project of the Ministry of Education in 2020: Research on the evaluation and promotion of teaching competency of new teachers in local universities in western China (EIA200358).

REFERENCES

[1] Kumar, P., Khokher, V., Gupta, Y., & Raman, B. (2021, September). Hybrid fusion based approach for multimodal emotion recognition with insufficient labeled data. In *2021 IEEE International Conference on Image Processing (ICIP)* (pp. 314-318). IEEE.

[2] Pandeya, Y. R., & Lee, J. (2021). Deep learning-based late fusion of multimodal information for emotion classification of music video. *Multimedia Tools and Applications*, *80*(2), 2887-2905.

[3] Zhang, Z., Wang, Z., Li, X., Liu, N., Guo, B., & Yu, Z. (2021). ModalNet: an aspect-level sentiment classification model by exploring multimodal data with fusion discriminant attentional network. *World Wide Web*, *24*, 1957-1974.

[4] Pawłowski, M., Wróblewska, A., & Sysko-Romańczuk, S. (2023). Effective techniques for multimodal data fusion: A comparative analysis. *Sensors*, *23*(5), 2381.

[5] Qi, Q., Lin, L., & Zhang, R. (2021). Feature extraction network with attention mechanism for data enhancement and recombination fusion for multimodal sentiment analysis. *Information*, *12*(9), 342.

[6] Zhang, K., Li, Y., Wang, J., Wang, Z., & Li, X. (2021). Feature fusion for multimodal emotion recognition based on deep canonical correlation analysis. *IEEE Signal Processing Letters*, *28*, 1898-1902.

[7] Tembhurne, J. V., & Diwan, T. (2021). Sentiment analysis in textual, visual and multimodal inputs using recurrent neural networks. *Multimedia Tools and Applications*, *80*(5), 6871-6910.

[8] Han, W., Chen, H., Gelbukh, A., Zadeh, A., Morency, L. P., & Poria, S. (2021, October). Bi-bimodal modality fusion for correlation-controlled multimodal sentiment analysis. In *Proceedings of the 2021 international conference on multimodal interaction* (pp. 6-15).

[9] Du, Y., Liu, Y., Peng, Z., & Jin, X. (2022). Gated attention fusion network for multimodal sentiment classification. *Knowledge-Based Systems*, *240*, 108107.

[10] Li, Z., Xu, B., Zhu, C., & Zhao, T. (2022). CLMLF: A contrastive learning and multi-layer fusion method for multimodal sentiment detection. *arXiv preprint arXiv:2204.05515*.

[11] Liu, W., Qiu, J. L., Zheng, W. L., & Lu, B. L. (2021). Comparing recognition performance and robustness of multimodal deep learning models for multimodal emotion recognition. *IEEE Transactions on Cognitive and Developmental Systems*, *14*(2), 715-729.

[12] Huddar, M. G., Sannakki, S. S., & Rajpurohit, V. S. (2021). Attention-based multimodal contextual fusion for sentiment and emotion classification using bidirectional LSTM. *Multimedia Tools and Applications*, *80*(9), 13059-13076.

[13] Cai, C., He, Y., Sun, L., Lian, Z., Liu, B., Tao, J., ... & Wang, K. (2021). Multimodal sentiment analysis based on recurrent neural network and multimodal attention. In *Proceedings of the 2nd on multimodal sentiment analysis challenge* (pp. 61-67).

[14] Peng, C., Zhang, C., Xue, X., Gao, J., Liang, H., & Niu, Z. (2021). Cross-modal complementary network with hierarchical fusion for multimodal sentiment classification. *Tsinghua Science and Technology*, *27*(4), 664-679.

[15] Chen, B., Cao, Q., Hou, M., Zhang, Z., Lu, G., & Zhang, D. (2021). Multimodal emotion recognition with temporal and semantic consistency. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *29*, 3592-3603.

[16] Luo, H., Ji, L., Huang, Y., Wang, B., Ji, S., & Li, T. (2021). Scalevlad: Improving multimodal sentiment analysis via multi-scale fusion of locally descriptors. *arXiv preprint arXiv:2112.01368*.

[17] Huddar, M. G., Sannakki, S. S., & Rajpurohit, V. S. (2020). Multi-level context extraction and attention-based contextual inter-modal fusion for multimodal sentiment analysis and emotion classification. *International Journal of Multimedia Information Retrieval*, *9*(2), 103-112.

[18] Huddar, M. G., Sannakki, S. S., & Rajpurohit, V. S. (2021). Attention-based multi-modal sentiment analysis and emotion detection in conversation using RNN.

[19] Sun, L., Xu, M., Lian, Z., Liu, B., Tao, J., Wang, M., & Cheng, Y. (2021). Multimodal emotion recognition and sentiment analysis via attention enhanced recurrent model. In *Proceedings of the 2nd on Multimodal Sentiment Analysis Challenge* (pp. 15-20).

[20] Tzirakis, P., Chen, J., Zafeiriou, S., & Schuller, B. (2021). End-to-end multimodal affect recognition in real-world environments. *Information Fusion*, *68*, 46-53.

[21] He, J., Yanga, H., Zhang, C., Chen, H., & Xua, Y. (2022). Dynamic invariant-specific representation fusion network for multimodal sentiment analysis. *Computational Intelligence and Neuroscience*, *2022*.

[22] Sun, J., Yin, H., Tian, Y., Wu, J., Shen, L., & Chen, L. (2021). Two-level multimodal fusion for sentiment analysis in public security. *Security and Communication Networks*, *2021*, 1-10.

[23] Zhang, J., Wu, X., & Huang, C. (2023). AdaMoW: multimodal sentiment analysis based on adaptive modality-specific weight fusion network. *IEEE Access*.

[24] Yan, X., Xue, H., Jiang, S., & Liu, Z. (2022). Multimodal sentiment analysis using multi-tensor fusion network with cross-modal modeling. *Applied Artificial Intelligence*, *36*(1), 2000688.

[25] Al-Tameemi, I. S., Feizi-Derakhshi, M. R., Pashazadeh, S., & Asadpour, M. (2023). Multi-model fusion framework using deep learning for visual-textual sentiment classification. *Computers, Materials & Continua*, *76*(2), 2145-2177.

[26] Younis, E. M., Zaki, S. M., Kanjo, E., & Houssein, E. H. (2022). Evaluating ensemble learning methods for multi-modal emotion recognition using sensor data fusion. *Sensors*, *22*(15), 5611.

[27] Njoku, J. N., Caliwag, A. C., Lim, W., Kim, S., Hwang, H., & Jung, J. (2022). Deep learning based data fusion methods for multimodal emotion recognition. *The Journal of Korean Institute of Communications and Information Sciences*, *47*(1), 79-87.

[28] Han, W., Chen, H., & Poria, S. (2021). Improving multimodal fusion with hierarchical mutual information maximization for multimodal sentiment analysis. *arXiv preprint arXiv:2109.00412*.

[29] https://www.kaggle.com/datasets/mathurinache/cmu-mosi

[30] Zhan, M., Zhao, C., Qin, K., Huang, P., Fang, M., & Zhao, C. (2023). Subaperture keystone transform matched filtering algorithm and its application for air moving target detection in an SBEWR system. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *16*, 2262-2274.

[31] Zhu, X., & Zheng, S. (2021). Uncertainty principles for the two-sided offset quaternion linear canonical transform. *Mathematical Methods in the Applied Sciences*, *44*(18), 14236-14255.

[32] Luo, Q., He, H., Liu, K., Yang, C., Silven, O., & Liu, L. (2023). Rain-like layer removal from hot-rolled steel strip based on attentive dual residual generative adversarial network. *IEEE Transactions on Instrumentation and Measurement*.

[33] Sadeeq, H. T., & Abdulazeez, A. M. (2022, September). Improved northern Goshawk optimization algorithm for global optimization. In *2022 4th International Conference on Advanced Science and Engineering (ICOASE)* (pp. 89-94). IEEE.