Deep Learning-Based Depression Analysis Among College Students Using Multi Modal Techniques

Liyan Wang
College of Education, Chuzhou City Vocational College, Chuzhou 239000, China

***Abstract*—This study proposed a novel approach to handle mental health, particularly, depression among college students, called CRADDS A Comprehensive Real-time Adaptive Depression Detection System. The novel CRADDS combined advanced tensor fusion networks which is able to analyze emotions using audio, text and video data more accurately, this is possible due to the strength of deep learning and multimodal approaches. This system is constructed with a hybrid algorithm framework that combines SVM (Support Vector Machines), CNN (Convolutional Neural Network) and (Bidirectional Long-Term Short-Term Memory) BiLSTM techniques. To address the limitations identified in earlier research, CRADDS increasing its feature set and using effective machine learning algorithms to reduce false positives and negatives. Further, it includes the advanced IoT devices to collect real time data from various range of public and private sources. The depression symptoms may be continuously monitored in real time, which helps to identify depressions in early stages and guaranteed the perfect well-being of students. Additionally, the model has the ability to adjust based on the interaction features, which helps to provide psychological support using the automatic responses observed from the verbal and nonverbal clues. Experiments show that the proposed CRADDS obtained an impressive accuracy based on the features of text, audio and video, when compared with the existing models. Overall, CRADDS is a useful tool for mental health professionals and educational institutions because it not only identifies depression but also helps to treat it earlier, and guarantees good academic scores and general well-being. The proposed validation accuracy increases from 63.04% to 86.08% which is higher than compared existing SVM model.**

***Keywords*—Depression analysis; multimodal techniques; mental health; real-time monitoring; hybrid algorithms**

I. Introduction

A. Depression Analysis and its Importance

Examining depression among students become very important, particularly in COVID-19 situations, which has severely increased mental health issues. Lockdowns and remote learning caused students to be away from their regular social networks and classrooms, which led to increased stress, anxiety [20] and depression symptoms in the students. Particular psychological difficulties were presented by the change to online learning environments, the disturbance of habits and future uncertainty [1]. The analysis of depression occurrence among students during this period was necessary to allow early detection and treatment, for preventing long-term mental health issues. By using effective depression analysis techniques, educational institutions and healthcare practitioners were able to develop and execute mental health interventions that were specifically designed to meet the needs of students who were experiencing difficulties during the pandemic [2-3]. These methods included wellness programs, peer support systems and online counselling services. Additionally, by understanding the patterns and situations regarding depression in students, educators and others can efficiently create academic and psychological support networks. COVID-19 raised focus to the importance of mental health measures in educational settings and highlighted the value of mental health as a fundamental element of overall well-being and successful learning [4-5]. In ongoing global health crisis, assessing student depression will provide valuable insights into the future approaches to student health services. It also highlights the importance of mental health plays in improving academic flexibility and success.

B. Depression Analysis Techniques and its Drawbacks

Depression analysis techniques involve a variety of methodologies, such as self-report surveys, clinician interviews and growing technology-based approaches like machine learning models that are used to analyse behavioural data [6]. Traditional self-report measures, like the Beck Depression Inventory and the Hamilton Depression Rating Scale, are commonly used, due to their adaptability and ability to track changes [7-8]. However, these previous literature methods can be unfair because sometimes people underestimate the symptoms due to the misunderstanding of questions. Observing nonverbal signals that indicate depression and further analysing patient responses can be done through clinician interviews, which provide a more understanding level of information [9-10]. Furthermore, using machine learning models provides an effective way to raise the accuracy of the depression diagnosis. These models may evaluate large amounts of data from various sources, like speech patterns, physical activity and social media usage, and identify patterns immediately that are not achievable with these traditional methods. Due to the limitations with these traditional techniques, there is an immediate need for multimodal based machine learning approaches. By analysing the advantages and trends of machine learning modals, we present the effective solution for this.

C. Machine Learning and its Advantages

Deep learning a subset of machine learning, provides a number of benefits when it comes to evaluating depression in college students by using advanced algorithms to understand a wide range of data sources. This technology is particularly good at immediate relationships and patterns that conventional analytical techniques could miss. For example, it can examine writing and speech patterns as well as social media activity to identify early indicators of depression that may not be immediately noticeable. Some of the techniques and its

advantages in reviewed by current research methods are illustrated below (Table I) [11-14].

The above researches obtain remarkable improvement in depression analysis with different data domains. Based on this research procedures, this study gives an advanced solution that tackle not only the present limitations but also the future.

Traditional depression analysis models frequently fail in numerous important domains when applied in real-time. Previous research mostly used discrete data modalities like text, audio, or video, which might result in assessments that are both incomplete and perhaps erroneous. These models often employ opaque black box techniques, which make it challenging to comprehend the decision-making process and pinpoint the fundamental causes of depression. Furthermore, the temporal dynamics and intricate connections included in multi-modal data pose challenges to the handling capabilities of many of the models that are now in use.

Our suggested CRADDS (Credit Risk Assessment Decision Support System) uses three potent algorithms—Convolutional Neural Network, BiLSTM, and SVM—to close these gaps. The individual qualities of each algorithm work together to improve the system's overall efficacy and accuracy in real-time depression analysis.

First, the CNN in CRADDS is enhanced with dilated convolutions, which increase the receptive field without compromising resolution, making it different from a standard convolutional network. This makes it possible for the model to extract more contextual information from the input photographs, which is important for detecting small changes in facial expressions and subtle emotional subtleties that could be signs of depression. Second, the model can concentrate on significant features from textual, audio, and video data sequences thanks to the attention mechanism built into the BiLSTM layer. This increases the model's capacity to represent intricate linkages and long-range dependencies, which raises the model's accuracy in identifying patterns of sadness over time. Finally, by combining visual, textual, and aural signals, SVM ensures robust categorization and greatly lowers the likelihood of false positives and negatives.

D. Proposed CRADDS Advantages and Study Motive

The propose study designed with an objective regarding three existing articles [15-17], limitations and future scope, this study not only focused on depression analysis, also provides an effective solution for the current research limitations, additionally, the future scope of the studies also completely satisfied with our proposed CRADDS. The possibility is clearly overviewed by Table II.

E. Depression Analysis among Various Factors

A thorough investigation of depression among medical students was carried out by Puthran et al. (2016), and the results showed that the frequency was 28.0% worldwide. Remarkably, the highest rates of depression were seen in Year 1 students, with a progressive drop noted in future years. Even if the rates of depression in medical and non-medical students were identical, the poor treatment behavior among depressed medical students highlights the need for targeted treatments. A comparatively high incidence of depression of 28.4% was carried out by Gao et al. (2020), which examined the prevalence of depression among Chinese university students. The subgroup analysis highlights the need for improved mental healthcare services for this and suggests a continuous requirement for interventions and support networks in Chinese colleges.

TABLE I. MACHINE LEARNING [21] TECHNIQUES AND ITS ADVANTAGES

Source	Techniques Used	Data Used	Improvements Noted
[11]	SVM, Naïve Bayes	Social Media Posts	Improved early detection accuracy
[12]	CNN, kNN, Random Forest	Facial Images, dynamic textual descriptions.	2.7% better in feature extraction.
[13]	Deep Learning, VGG-16, Word2Vec, Faster R-CNN	Social Media Posts (texts, images, videos)	First real-time multimodal analysis system.
[14]	BiLSTM	Textual posts on social media	Good results in early depression detection.

TABLE II. LIMITATIONS AND FUTURE SCOPE OF EXISTING RESEARCH

Source	Limitations	Future Scope	How CRADDS address Limitations and Future Scope
[15]	High risk of false positives and negatives, Ethical concerns	Expand the use of IoT for real-time diagnostics Integrate with voice conversation systems for therapeutic effects	Implements robust validation to minimize diagnostic errors Designs ethical AI frameworks and observes to guidelines Improves IoT integration and supports real-time multimodal analysis
[16]	Relies on audio and text; plans for video integration Requires broader, more accurate datasets	Develop a hybrid model using audio, video, and text features Implement more powerful algorithms for enhanced accuracy	Uses a comprehensive multimodal approach integrating audio, text, and video Applies advanced algorithms to improve learning rates and prediction accuracy Plans for real-time, scalable depression detection applications
[17]	Limited participant number affects result validity Manual collection of verbal and non-verbal cues is resource-intensive	Develop automatic monitoring through app Use advanced statistical analysis for more significant findings Reduce required data collection period	Expands dataset to include more demographic variables for greater representativeness Combines automatic monitoring of verbal and non-verbal cues through mobile apps Applies machine learning to reduce data collection period while maintaining accuracy

Machine learning approaches were used by Qasrawi et al., (2022) to predict risk factors related to anxiety and depression in school-age children. The models with the best accuracy levels were SVM and RF, underscoring the importance of variables including family income, academic performance, home environment and violence in schools in impacting mental health symptoms. The results suggest that to improve mental health preventive and intervention programs, machine learning should be included into school information systems. Haque et al., (2021) used machine learning techniques to identify depression in kids and teens between the ages of 4 and 17. After predicting depressed classes with a high accuracy rate of 95%, RF was shown to be the most effective algorithm. Suicidal thoughts, sleep difficulties, and mood-related symptoms were important indicators of depression, highlighting the need of early identification and treatment to lessen the harmful impacts of depression in this susceptible group.

The remaining sections of the article are discussed in four sections. In Section II methods of the proposed model are outlined. In Section III, the results of the experiments are discussed. In Section IV, the conclusion is presented.

II. METHOD

A. Proposed Model Outline

The foundation of our proposed CRADDS is the combination of three powerful algorithms: SVM (Support Vector Machine), CNN (Convolutional Neural Network), and BiLSTM (Bidirectional Long Short-Term Memory). Each of these algorithms includes specific features to improve the system's effectiveness and precision in real-time depression analysis.

1) CNN: CRADDS's CNN is not like a regular convolutional network; it is improved by convolutional layers with specific functions that make use of dilated convolutions. These dilated convolutions increase the network's sensitive field without sacrificing resolution, allowing the model to extract more contextual information from input images. This is important for identifying detail emotions in recognition tasks. This is especially important for identifying changes in video expressions that could point to despair.

2) Bi-LSTM: Bi-LSTM layer of CRADDS is used to give importance to certain data points. Its attention-mechanism allows the algorithm to focus more on important features from textual, audio and video data sequences that have a better ability to identify depression. The model's ability to learn from difficult dependencies and long-range connections in the data, which is made possible by weighting input information differently and improves its ability to observe depression patterns in time.

3) SVM: Together with these advanced techniques of CNN and Bi-LSTM, SVM strength also added to make CRADDS effective. To conduct detailed analysis, the system continuously combines visual, textual and audio signals and greatly reduce the possibility of false positives and negatives. Through the combination of these advanced algorithms, CRADDS improve diagnostic precision and acts as an effective tool for early identification of depression, and guaranteeing quick support for depressed individuals.

B. Architecture

1) CNN architecture: In this section the proposed CRADDS used a dilated convolutional neural network (DCNN) to analyse depression very accurately. Because the dilated kernel is a perfect tool to analyse depression in any form of audio, video and textual. DCNN is important for improving the ability to analyse difficult emotional signals from multiple methods such as speech patterns, facial expressions, and textual data words. Traditional convolutional kernels are defined by

$$ot_w = \left(\frac{it_w - n + 2p}{s}\right) + 1 \quad (1)$$

$$ot_h = \left(\frac{it_h - n + 2p}{s}\right) + 1 \quad (2)$$

ot_w and ot_h are the output width and height respectively. it_w and it_h are the input height and width. n denotes the size of convolutional filter and p is the amount of padding applied to the input. s is the stride which the kernel moves across the input. The concept of traditional techniques is updated by using dilated convolutions which is used to extract the input features under CRADDS. d is the dilation factor. By introducing gaps into the kernel, dilation allows the network to have a bigger responsive field by effectively raising the kernel size without increasing the number of weights.

$$ot_w = \left(\frac{it_w - (n-1)\times(d-1) + 2p}{s}\right) + 1 \quad (3)$$

$$ot_h = \left(\frac{it_h - (n-1)\times(d-1) + 2p}{s}\right) + 1 \quad (4)$$

Here d is the dilation rate. $(n-1)\ and\ (d-1)$ adjusts the kernel size by considering the gaps inserted between the kernel's elements to modify the kernel's size. In CRADDS, we build the DCNN model by replacing these with dilated convolution kernels. By adding gaps to the kernel grid, dilated convolutions increase the field of contact without adding to the computational complexity. For example, the receiving area effectively grows from 3x3 to 7x7 and, with further dilation, to 15x15 by changing conventional 3x3 kernels to include dilations. Even with these increases, the total number of parameters stays fixed, preventing higher processing expenses and improving the network's ability to extract more detailed information from the input data.

Using a range of dilation rates that are carefully selected to capture the serious patterns related to emotion changes and emotional states in depression, the DCNN processing is improved for the identification of depression. Each of the dilation rates 1, 2 and 4 is precisely adjusted to the feature scales that are important for emotional analysis. The softmax function is defined as

$$\sigma(zj) = \frac{e^{zj}}{\sum_{j-1}^{J} e^{zj}} \quad (5)$$

In Eq. (5), zj denotes the element in vector z with j highlights the total number of elements. Several dilation rates are built into the architecture of the DCNN in CRADDS, which improves feature extraction abilities and guaranteeing full

coverage of the input data. 6 dilated convolution-pooling modules, two fully connected layers, and a softmax output layer make up the DCNN structure. Dropout functions are integrated to reduce overloading, maintain the integrity of input information and improve performance. The specified dilations are defined as

$$mi = \max[m(i+1) - 2ri, m(i+1) - 2(m(i+1) - ri), ri] \quad (6)$$

Here mi is the dilation rate for the current layer (i), $m(i+1)$ is the dilation rate for the next layer $(i+1)$ and ri is a parameter. The structure of DCNN is visually presented under Fig. 1.

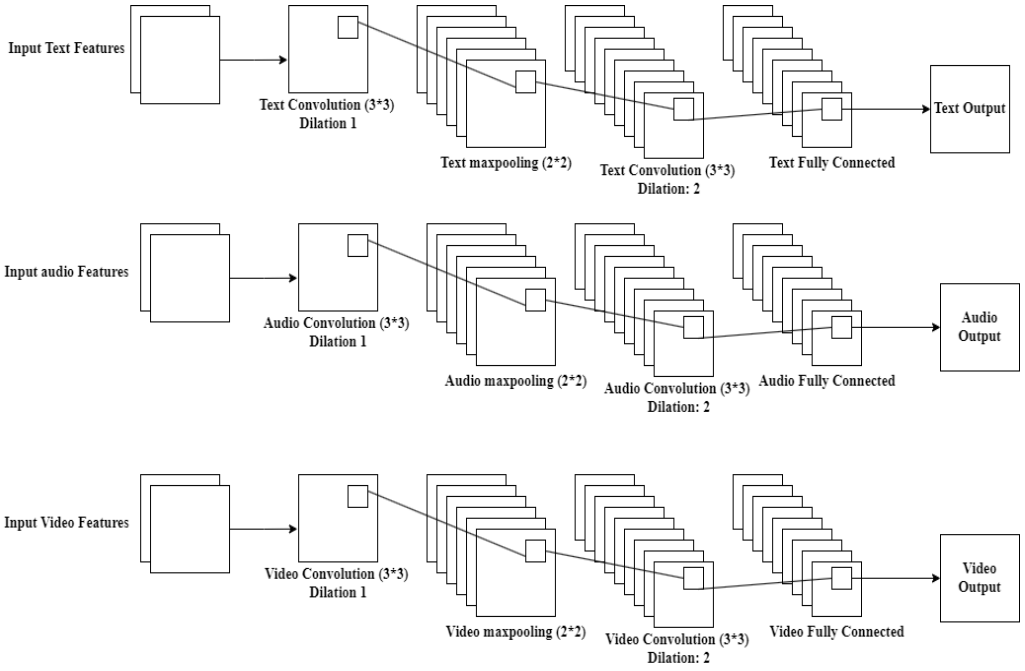

Fig. 1. DCNN structure for depression analysis for text, audio and video data.

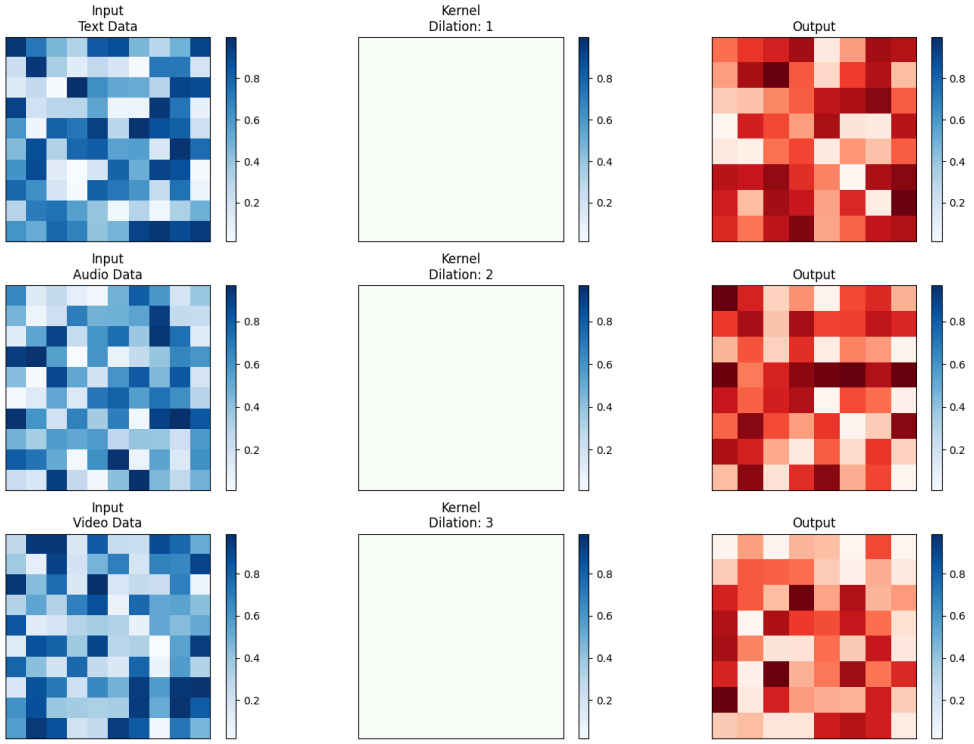

Fig. 2. Dilation results for text, audio and video data under CRADDS.

Fig. 2 shows the exact dilation process of text, audio and video inputs. For text processing, the kernel has been set for textual input with a dilation rate of 1, which indicates a conventional convolution that is direct and does not have any gaps. When analysing text, local information such as associations between words are important for understanding emotions. This minimal dilation is suitable for text processing. For audio processing, figure displays a kernel with a dilation rate of 2 for audio data. The kernel covers a greater portion of the input due to the higher dilation, ignoring some data points in order to capture more extensive temporal patterns in the spectrogram, such as changes over time that are important for audio analysis. For features like pitch and tone that change over a series of samples, this type of dilation is useful for detecting patterns across somewhat longer time spans. Dilation rate of 3 is used to denote the video data processing, allows the convolutional process to cover a larger region of the input frames. This method works well with videos, because it can able to capture spatial relationships in larger regions, which is useful when detecting movements and changes in videos by using many pixels to present the movements with high accuracy. By increasing dilation rate, the network will improve the area where it receives and include more related information from the video frames. This can be used to understand the challenging patterns in motion tasks and improve the accuracy to find out emotional expressions very clearly.

2) Bi-LSTM: CRADDS used BiLSTM with attention mechanism; by using its advanced features, it helps to improve the understanding of text, audio and video input. This model aims to identify the temporal patterns that are important for identifying depressions very accurately. Bi-LSTM layers allow the network to learn from data in both forward and backward directions. This helps the network to capture the various temporal features effectively than the traditional LSTM. This bidirectional learning is important to CRADDS because it obtains a thorough understanding of the data, which can be the textual, audio and video clippings. Thus, the attention techniques used in BiLSTM highlights the particular data in to segments that are helpful to identify depression. The attention mechanism is expressed as

$$\begin{cases} ot, h = BiLSTM\ (a) \\ ot = [ot_f, ot_b] \\ ot = ot_f + ot_b \\ \omega = w \times ot + b \\ c = \tanh(ot) \times \omega \\ y = ot \times c \end{cases} \quad (7)$$

In Eq. (7), the inputs are denoted by a, the forward and backward LSTM outputs are represented by ot_f, ot_b , respectively, and their concatenation output is represented by ot. The weight vector ω and the weighted context c improve the model's ability to observe significant depression indications by focusing its learning on the most crucial elements of the sequence.

The fully connected (FC) network processes the processed features after the attention layer, combining them into a final output that can be used to identify the presence and severity of depression. With this setup, each modality of text, audio and video is evaluated separately and their insights are integrated to create a more accurate evaluation. Table III shows the parameter setting of the proposed Bi-LSTM.

TABLE III. PARAMETER SETTING OF PROPOSED BI-LSTM

Input Type	Layer Name	Parameter Setting
Text	Bi-LSTM	Hidden Units 128
	Layer	Layers 2
		Dropout 0.5
	Attention	Dropout 0.5
	FC1	Output Features 128
		ReLU
		Dropout 0.5
	FC2	Output Features 128
		ReLU
Audio	Bi-LSTM	Hidden Units 128
	Layer	Layers 2
		Dropout 0.5
	Attention	Dropout 0.5
	FC1	Output Features 128
		ReLU
		Dropout 0.5
	FC2	Output Features 128
		ReLU
Video	Bi-LSTM	Hidden Units 128
	Layer	Layers 2
		Dropout 0.5
	Attention	Dropout 0.5
	FC1	Output Features 128
		ReLU
		Dropout 0.5
	FC2	Output Features 128
		ReLU

3) Multi-modal fusion: Additionally, embeddings from the last Bi-LSTM layer and a DCNN processing features are concatenated to address the multimodal character of the input. By feeding this concatenated vector into a further FC layer, the results obtained from the analysis of text, audio and video are successfully combined.

$$fo_t, x_{ba_{fused}} = [DCNN\ (a_{txt}), DCNN\ (a_{audio}), DCNN\ (a_{video}) \quad (8)$$

Here fi_t denotes fused input of DCNN text, audio and video outputs respectively.

BiLSTM processing of concatenated features

$$y_{temp} = BiLSTM\ (x_{ba_{fused}}) \quad (9)$$

Here $(x_{ba_{fused}})$ is the concatenated vector from all three modalities after initial DCNN processing. y_{temp} denotes the output from Bi-LTSM which produces temporal and sequential information across the multimodal data. The final prediction is expressed as

$$y_{pred} = FC(w_{fuse} * y_{temp} + b_{fuse}) \quad (10)$$

In Eq. (10) FC denotes fully connected network that combines the multimodal temporal features into final predictive output. w_{fuse} and b_{fuse} denotes weights and biases of the final FC layer.

To improve the system the loss function needs to consider the combined influence of text, audio and video data. This can be expressed as,

$$L = \ell(y_{pred}, y) \tag{11}$$

ℓ is the chosen loss function, cross entropy for classification tasks.

4) SVM based feature extraction: The SVM is mainly used for feature extraction from difficult, high-dimensional datasets in our proposed CRADDS study. To improve the margin between two classes, the initial stage in this approach is to define a separating hyperplane using the traditional SVM technique for supervised learning classification. This can be expressed as

$$\min \frac{1}{2}\ ||W||^2 + C\sum_{i=1}^{N}\xi_i \text{ subject to}$$

$$ti(W^T Xi + B) \geq 1 - \xi_i, \xi_i \geq 0, \quad i = 1, \ldots., N \tag{12}$$

Where, the balance between increasing the margin and reducing classification mistakes is expressed by C, and ξ_i are slack variables that account for misclassifications. By applying higher boundaries and promoting accurate classification, a high C value helps to reduce misclassification.

The SVM successfully uses the kernel method to handle the non-linear aspects of energy system data. RBF (Radial Basis Function) kernel is expressed as,

$$K(Xi, Xj) = e^{-\frac{1}{2\sigma^2}||Xi - Xj||^2} \tag{13}$$

where the flexibility of the kernel function is controlled by the kernel parameter σ^2. The SVM can operate in a converted feature space where non-linear connections are corrected, allowing the separation of data points that are not linearly separable in the original space. This kernel simplifies this process. In addition, we incorporate a cost matrix into the SVM to handle the issues arising from dataset imbalances, which might lead to bias in the classification boundaries in favour of the majority class. This matrix reduces bias by adjusting the misclassification penalty to prioritize the minority class. The cost matrix function expressed as

$$co = \begin{bmatrix} 0 & 1 \\ c & 0 \end{bmatrix} \tag{14}$$

if $c > 1$, then it would cost more to incorrectly classify an instance of the minority class than the majority class. This strategy gives a more equitable categorization result by bringing the boundary closer to the majority class, which makes the model more sensitive to the minority class. The model reduces dimensionality and separates the essential elements from the input energy data through this procedure, guaranteeing reliable prediction outcomes. Fig. 3(a), 3(b) and 3(c) present the process of SVM classification of text, audio and video input.

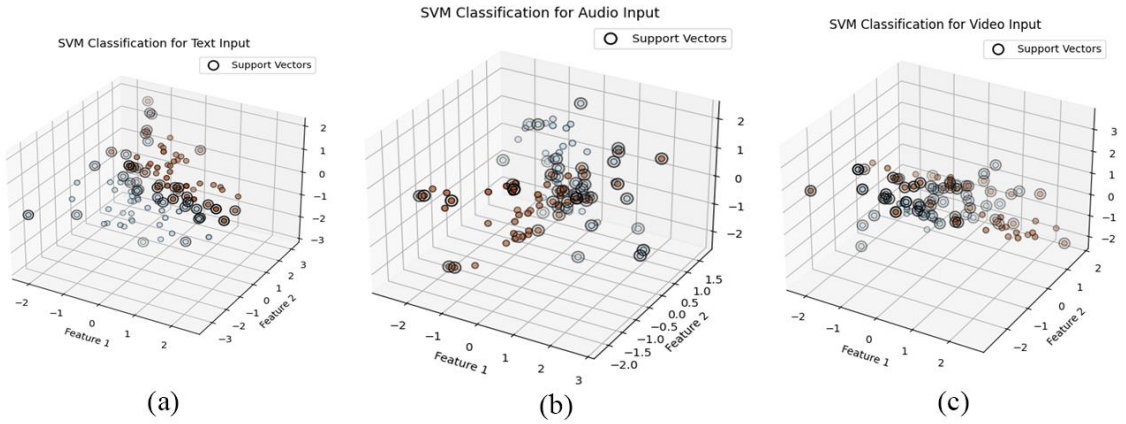

Fig. 3. SVM classification on Text input, Audio input, and Video input.

III. Results and Experiments

A. Simulation Setup

Proposed CRADDS is evaluated using DAIC-WOZ datasets adapted from [16]. Based on that Table IV presents the features of dataset which is used to evaluate proposed CRADDS.

B. Evaluation Criteria

In the present study, the results of the CRADDS are compared with the three existing researches of [15] [16] [17]. The main objective of the CRADDS is to address the limitation of these studies and also satisfy the future visions. Based on the task we proceed with an experiment.

Table V presents that the CRADDS model performs significantly well when tested on text, audio and video data using DCNN, BiLSTM, and SVM. The validation accuracy and loss for Text DCNN are 0.45 and 0.85, respectively, and the training accuracy is 0.94 with a loss of 0.25. Using validation metrics of 0.82 accuracy and 0.28 loss, Audio DCNN achieves a training accuracy of 0.96 with a reduced loss of 0.12. Video DCNN validation accuracy of 0.83, a validation loss of 0.35, and a training accuracy of 0.95 and loss of 0.22. The validation accuracy and loss for Text BiLSTM are 0.80 and 0.30, and the accuracy is 0.89 with a loss of 0.18. With validation metrics of 0.81 accuracy and 0.25 loss, Audio BiLSTM exhibits 0.91 accuracy and 0.15 loss. With a validation accuracy and loss of 0.80 and 0.28, Video BiLSTM exhibits an accuracy of 0.90 and

a loss of 0.17. Text SVM achieves validation accuracy of 0.78 and loss of 0.32, together with training accuracy of 0.88 and 0.20 loss. Audio SVM records validation accuracy and loss of 0.79 and 0.27, along with 0.92 training accuracy and 0.14 loss. Lastly, Video SVM displays validation accuracy and loss of 0.77 and 0.30 with 0.90 training accuracy and 0.19 loss. These findings show that, for all data types, DCNN models perform more accurately than BiLSTM and SVM, with Audio DCNN shows the best overall performance.

TABLE IV. DATASET FEATURES

Category	Description	Category	Description
Dataset	DAIC-WOZ Depression Database	**Participants**	59 Depressed; 130 non-depressed individuals
Purpose	Automatic Depression Detection System	**Data Types**	Audio recordings (AUDIO.wav) Video recording Text responses (TRANSCRIPT.csv, FORMANT.csv, etc.)
Source	University of Southern California (USC)	**Training Set**	IDs of patients Patient PHQ-8 scores Binary labels Gender Questionnaire responses
Access	Apply on USC website for access and download	**Development Set**	IDs of patients Patient PHQ-8 scores Gender Binary labels Questionnaire responses
Data Format	Zip files (189 sessions: from 300 P.zip to 492 P.zip)	**Test Set**	IDs of patients Gender
Total Sessions	189	**Features**	Verbal symptoms Non-verbal symptoms Audio features Video features Text features

TABLE V. EVALUATION PARAMETERS FOR PROPOSED CRADDS

Method	Tra-Accuracy	Tra-Loss	Val-Accuracy	Val-Loss
Text DCNN	0.94	0.25	0.85	0.45
Audio DCNN	0.96	0.12	0.82	0.28
Video DCNN	0.95	0.22	0.83	0.35
Text BiLSTM	0.89	0.18	0.80	0.30
Audio BiLSTM	0.91	0.15	0.81	0.25
Video BiLSTM	0.90	0.17	0.80	0.28
Text SVM	0.88	0.20	0.78	0.32
Audio SVM	0.92	0.14	0.79	0.27
Video SVM	0.90	0.19	0.77	0.30

TABLE VI. PERFORMANCE EVALUATION OF PROPOSED CRADDS

Method	Precision	Recall	F1	Support
Text DCNN	0.93	0.92	0.93	50
Audio DCNN	0.93	0.90	0.91	50
Video DCNN	0.93	0.90	0.87	50
Text BiLSTM	0.82	0.85	0.83	50
Audio BiLSTM	0.84	0.86	0.85	50
Video BiLSTM	0.83	0.85	0.84	50
Text SVM	0.80	0.82	0.81	50
Audio SVM	0.82	0.84	0.83	50
Video SVM	0.82	0.83	0.83	50

C. Performance Comparison with Existing Studies

As we discussed earlier, in this section the proposed CRADDS based techniques of DCNN, BiLSTM with attention mechanism and SVM are compared with the existing research studies of [15] [16] and [17].

Fig. 4 presents the efficacy of CRADDS based DCNN when compared with the efficacy of CNN [15]. The performance of the DCNN-based CRADDS on training and validation datasets obtains a notable efficacy in the depression diagnosis. The model's ability to adapt to new data is confirmed by the figure, which shows how training and validation loss meet. The validation loss decreases from 18 to 2.5 while the training loss drops substantially from 20 to 1.5 during the epochs, demonstrating the model's capacity for learning and error reduction. At the same time, the training accuracy steadily increases to 95.03%, whereas the validation accuracy rises steadily to 82.10%. These show that multimodal data including text, audio and video inputs has complex patterns that the DCNN is able to capture successfully. Comparative studies indicate that the model outperforms typical CNN models in reliably identifying depression, as seen by its higher precision and recall. Table VI shows performance evaluation of proposed CRADDS.

Fig. 5 shows, when comparing the CNN-LSTM model [16] to the proposed CRADDS model BiLSTM, it shows remarkable efficacy in depression diagnosis. The training loss decreased from 18 to 2 and the validation loss from 16 to 3, respectively, on the training and validation loss, which show a considerable reduction across epochs. The immediate drop in loss values presents how well the BiLSTM model learns and adapt from the data. The validation accuracy increases gradually to

85.04%, but the training accuracy curve shows a continuous improvement up to 94.07%. These findings highlight the BiLSTM capacity to efficiently extract difficult patterns and temporal connections from multimodal data that includes text, audio and video inputs. The BiLSTM in CRADDS shows better performance than the CNN-LSTM model, which is important for depression identification. CRADDS with BiLSTM is an effective tool for automatic depression identification because of its improved feature extraction and classification abilities.

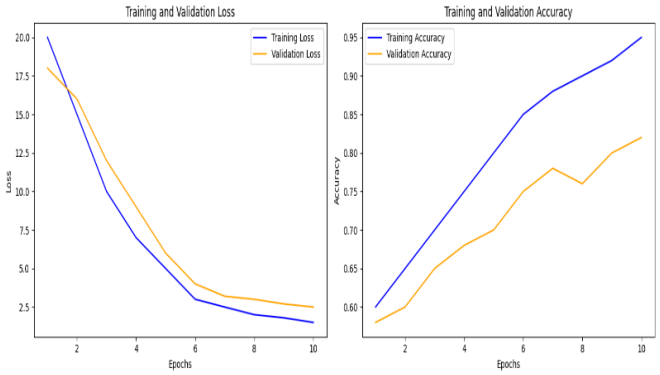

Fig. 4. CRADDS-based DCNN results against typical CNN [15] over Epochs.

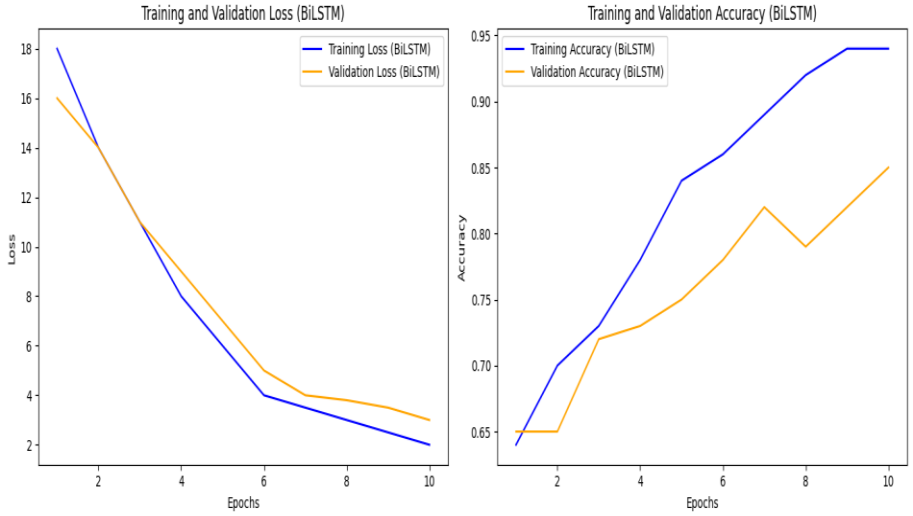

Fig. 5. CRADDS-based BiLSTM results against CNN-LSTM [16] over Epochs.

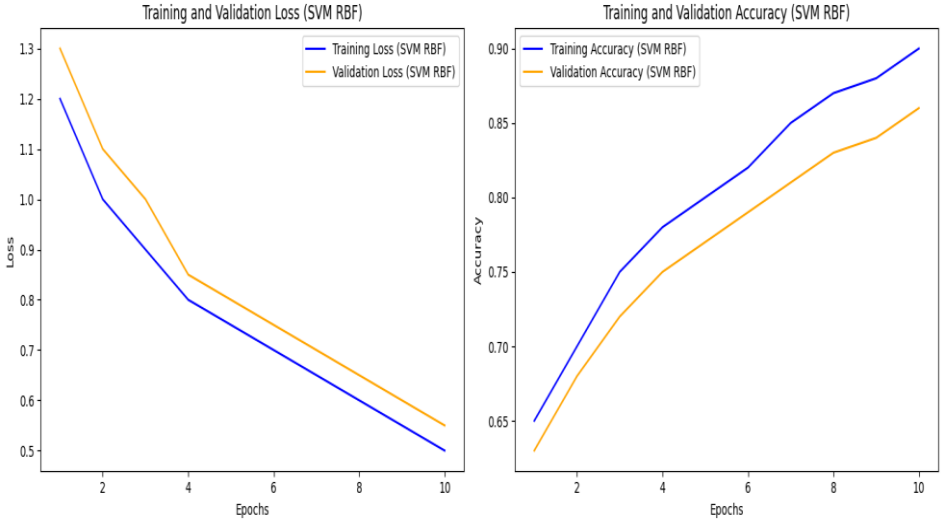

Fig. 6. CRADDS-based SVM(RBF) results against SVM [17] over Epochs.

Fig. 6 shows the efficacy of proposed CRADDS based SVM, when compared to the SVM model of [17], shows a notable improvement in depression identification [18, 19]. Over the course of the epochs, the training and validation loss figures show a constant decrease: the training loss dropped from 1.2 to 0.5 and the validation loss from 1.3 to 0.55. This steady decrease shows how well the model can adapt to new data. There is a consistent improvement in training accuracy from 65.12% to 90.02% and in validation accuracy from 63.04% to 86.08%. These show the effectiveness of SVM model learns and captures the difficult correlations found in the multimodal data (text, audio, and video). The CRADDS-based SVM model appears to be more effective at differentiating between people who are depressed and those who are not, based on its greater accuracy and lower loss values when compared to the regular SVM model.

IV. Conclusion

The study introduces a novel CRADDS system to analyse the depression among college students by using their posts regarding text, audio and video inputs under the platform of University of Southern California (USC) by using DAIC-WOC dataset. The proposed CRADDS uses the techniques of DCNN, BiLSTM and SVM (RBF Kernel) model. This study presents the unique objectives in the domain of depression analysis. In a modern day the techniques of deep learning are mostly used under wide range of applications, this study also uses the effective fusion techniques of deep learning algorithms. To make sure about the effectiveness of proposed CRADDS each technique of CRADDS is evaluated and compared against the existing effective techniques analysed form the study [15] [16] and [17]. The main motive of the present study is to address the limitation of these existing researches and to satisfy their future scope expectations. The proposed CRADDS have the ability to address these objectives which is discussed earlier under the Table II. The effective experiments regarding the Table II are demonstrated under Section IV. The results of proposed CRADDS highlights that the techniques of CRADDS based DCNN, BiLSTM and SVM are outperforms with their proposed techniques of the existing studies with their remarkable scores. The output obtained from all the models under CRADDS highlights its efficacy regarding the input features of text, audio and video format. Overall, the proposed achieves the best solution when compared with the existing studies objective and acts as an effective tool to meet not only the present but also the future demands under the investigation of depression, guaranteeing the perfect well-being of students as well as common individuals.

In order to improve the accuracy and robustness of the model, future study will investigate the integration of new data modalities, such as physiological signals. Our goal is to create edge computing-based real-time deployment solutions that increase efficiency and accessibility. Furthermore, investigating explainable AI methods will aid in improving the transparency and comprehensibility of the model's judgments. Finally, adding more demographic groupings to the dataset will guarantee the model's wider applicability and fairness.

Acknowledgment

This research was supported by Anhui Provincial Department of Education University Outstanding Young Talents Support Program (gxyqZD2020140) and Anhui Provincial Education Department Provincial Quality Engineering Project (2021zyjxzyk024).

References

[1] J. Bueno-Notivol, P. Gracia-García, B. Olaya, I. Lasheras, R. López-Antón, J. Santabárbara, “Prevalence of depression during the COVID-19 outbreak: A meta-analysis of community-based studies,” International Journal of Clinical and Health Psychology, vol. 21, no. 1, pp.100196, January-April 2021.

[2] M. G. Mazza, R. De Lorenzo, C. Conte, S. Poletti, B. Vai, I. Bollettini, E. M. T. Melloni, R. Furlan, F. Ciceri, P. Rovere-Querini, F. Benedetti, “Anxiety and depression in COVID-19 survivors: Role of inflammatory and clinical predictors,” Brain, Behavior, and Immunity, vol. 89, pp. 594-600, July 2020.

[3] J. Deng, F. Zhou, W. Hou, Z. Silver, C. Y. Wong, O. Chang, E. Huang, Q. K. Zuo, “The prevalence of depression, anxiety, and sleep disturbances in COVID-19 patients: a meta-analysis,” Annals of the New York Academy of Sciences, vol. 1486, no. 1, pp. 90-111, February 2021.

[4] Sommerlad, L. Marston, J. Huntley, G. Livingston, G. Lewis, A. Steptoe, D. Fancourt, “Social relationships and depression during the COVID-19 lockdown: longitudinal analysis of the COVID-19 Social Study,” Psychological Medicine, vol. 52, no. 15, pp. 3381-3390, January 2022.

[5] J. H. Lee, H. Lee, J. E. Kim, S. J. Moon, E. W. Nam, “Analysis of personal and national factors that influence depression in individuals during the COVID-19 pandemic: a web-based cross-sectional survey,” Globalization and Health, vol. 17, pp. 1-12, January 2021.

[6] P. D. Barua, J. Vicnesh, O. S. Lih, E. E. Palmer, T. Yamakawa, M. Kobayashi, U. R. Acharya, “Artificial intelligence assisted tools for the detection of anxiety and depression leading to suicidal ideation in adolescents: a review,” Cognitive Neurodynamics, vol. 18, no. 1, pp. 1-22, November 2022.

[7] Hajduska-Dér, G. Kiss, D. Sztahó, K. Vicsi, L. Simon, “The applicability of the Beck Depression Inventory and Hamilton Depression Scale in the automatic recognition of depression based on speech signal processing,” Frontiers in Psychiatry, vol. 13, pp. 879896, August 2022.

[8] Y. P. Wang, C. Gorenstein, “Assessment of depression in medical patients: a systematic review of the utility of the Beck Depression Inventory-II,” Clinics, vol. 68, pp. 1274-1287, September 2013.

[9] Nickel, G. Thomalla, “Post-stroke depression: impact of lesion location and methodological limitations—a topical review,” Frontiers in neurology, vol. 8, pp. 291355, September 2017.

[10] H. Byeon, “Advances in machine learning and explainable artificial intelligence for depression prediction,” International Journal of Advanced Computer Science and Applications, vol. 14, no. 6, pp. 520-526, July 2023.

[11] S. Smys, J. S. Raj, “Analysis of deep learning techniques for early detection of depression on social media network-a comparative study,” Journal of trends in Computer Science and Smart technology (TCSST), vol. 3, no. 1, pp. 24-39, 2021.

[12] P. Meshram, R. K. Rambola, “Diagnosis of depression level using multimodal approaches using deep learning techniques with multiple selective features,” Expert Systems, vol. 40, no. 4, pp. e12933, January 2023.

[13] Malhotra, R. Jindal, “Multimodal deep learning-based framework for detecting depression and suicidal behaviour by affective analysis of social media posts,” EAI Endorsed Transactions on Pervasive Health and Technology, vol. 6, no. 21, January 2020.

[14] F. M. Shah, F. Ahmed, S. K. S. Joy, S. Ahmed, S. Sadek, R. Shil, M. H. Kabir, “Early depression detection from social network using deep learning techniques,” In 2020 IEEE Region 10 Symposium (TENSYMP), pp. 823-826, June 2020.

[15] H. Yoo, H. Oh, "Depression detection model using multimodal deep learning," Preprints, May 2023.

[16] N. Marriwala, D. Chaudhary, "A hybrid model for depression detection using deep learning," Measurement: Sensors, vol. 25, pp. 100587, February 2023.

[17] R. P. Thati, A. S. Dhadwal, P. Kumar, P. Sainaba, "A novel multi-modal depression detection approach based on mobile crowd sensing and task-based mechanisms," Multimedia Tools and Applications, vol. 82, no. 4, pp. 4787-4820, April 2022.

[18] R. Puthran, M. W. Zhang, W. W. Tam, R. C. Ho, "Prevalence of depression amongst medical students: A meta-analysis," Medical Education, vol. 50, no. 4, pp. 456-468, March 2016.

[19] L. Gao, Y. Xie, C. Jia, W. Wang, "Prevalence of depression among Chinese university students: a systematic review and meta-analysis," Scientific reports, vol. 10, no. 1, pp. 15897, September 2020.

[20] R. Qasrawi, S. P. V. Polo, D. A. Al-Halawa, S. Hallaq, Z. Abdeen, "Assessment and prediction of depression and anxiety risk factors in schoolchildren: machine learning techniques performance analysis," JMIR formative research, vol. 6, no. 8, pp. e32736, August 2022.

[21] U. M. Haque, E. Kabir, R. Khanam, "Detection of child depression using machine learning methods," PLoS One, vol. 16, no. 12, pp. e0261131, December 2021.