https://doi.org/10.48047/AFJBS.6.14.2024.7389-7425

African Journal of Biological Sciences

Journal homepage: http://www.afjbs.com

Research Paper Open Access

Comparative Study of current methods and trends in AI to evaluate mental health using sentiment analysis

Megha Agarwal[1] , Dr. Vinodini Katiyar[2], Dr. Vandana Patel[2], Dr. Dinesh Kumar Singh[2]

[1] Research Scholar, Department of IT, Dr. Shakuntala Misra National Rehabilitation University, UP-India

[2]Head-Department of IT, Dr. Shakuntala Misra National Rehabilitation University, UP-India

[2]Professor, Department of Applied Sciences, B.N.College of Engineering and Technology, UP-India

[2]Assistant Professor-Department of IT, Dr. Shakuntala Misra National Rehabilitation University, UP-India

meghaagarwal2011@gmail.com

Volume 6, Issue 14, Aug 2024

Received: 15 June 2024

Accepted: 25 July 2024

Published: 15 Aug 2024

doi: 10.48047/AFJBS.6.14.2024.7389-7425

Abstract

Depression is a widespread mental health condition affecting millions globally, making early detection and intervention critical for effective treatment and management. With the proliferation of social media platforms, there is increasing interest in utilizing these platforms to identify signs of depression in users. This paper reviews current methods and trends in detecting depression through social media, highlighting key techniques, challenges, and future directions.

Depression is a prevalent mental disorder that significantly impacts individuals' mental health and daily lives. It often leads to a loss of interest in regular activities and can result in suicidal thoughts, making it a significant societal issue. Consequently, there is a growing need for automated systems to detect depression across various age groups. Researchers have been exploring effective approaches to identify depression, leading to numerous proposed studies.

This study analyzes existing research on the application of Artificial Intelligence (AI) and Machine Learning (ML) techniques for detecting depression. It examines different methods used to detect emotions and moods in individuals, including facial expressions, images, emotional chatbots, and texts on social media platforms. Techniques such as Naive-Bayes, Support Vector Machines (SVM), Long Short-Term Memory (LSTM) networks, Radial Neural Networks (RNN), Logistic Regression, and Linear Support Vector Machines are utilized to recognize emotions from text processing. Additionally, Artificial Neural Networks (ANN) are employed for feature extraction and classification of images to detect emotions through facial expressions.

This paper aims to survey various AI and ML techniques that assist in the detection and analysis of emotions, thereby identifying depression. It also addresses related research issues, providing a comprehensive overview of the state-of-the-art in this field.

1. Introduction

Depression, marked by persistent sadness and loss of interest, is a significant public health issue with both emotional and physical symptoms. Traditional diagnosis methods, like self-reported questionnaires and clinical interviews, face limitations due to stigma and accessibility. Social media offers a new avenue to detect depression by analyzing users' online behavior and language. In our fast-paced society, depression often stems from the inability to manage evolving societal pressures. Initial visual signs of depression, such as bags under the eyes, dried skin, and facial rashes, are detectable. Researchers have developed image processing algorithms to observe these facial changes and aid in diagnosis.

Sentiment analysis (SA) has garnered significant attention in artificial intelligence (AI) and natural language processing (NLP) due to the growing demand for automated analysis of user sentiment towards products or services. As opinions are increasingly shared online in videos rather than text, Multimodal Sentiment Analysis (MSA), which uses multiple modalities, has become an important research area. MSA leverages advancements in machine learning and deep learning for multimodal feature extraction, fusion, and sentiment polarity detection, aiming to minimize error rates and improve performance.

Recent MSA architecture developments are categorized into early fusion, late fusion, hybrid fusion, model-level fusion, tensor fusion, hierarchical fusion, bi-modal fusion, attention-based fusion, quantum-based fusion, and word-level fusion. These architectures have been compared in terms of their strengths and limitations. Social media generates vast amounts of unstructured information, requiring business organizations to process and analyze sentiments for insights. Historically, various machine learning and NLP approaches have been employed for sentiment analysis, but deep learning methods have gained popularity for their high performance.

Key contributions from researchers focus on deep learning approaches, which are highlighted in sentiment analysis tasks across multiple languages. Deep learning has emerged as a powerful technique for multimodal sentiment analysis, prompting the need for survey papers summarizing recent trends. This synopsis categorizes thirty-five state-of-the-art models in video sentiment analysis into eight categories based on their architectures. The effectiveness of these models is evaluated using the CMU-MOSI and CMU-MOSEI datasets.

The Multi-Modal Multi-Utterance based architecture has been identified as the most powerful for multimodal sentiment analysis, utilizing information from all modalities and contextual information from neighboring video utterances. This architecture includes a Context Extraction Module, often a bidirectional recurrent neural network, and an Attention-Based Module, which fuses text, audio, and video modalities, prioritizing the most important ones. These findings aim to provide newcomers with a comprehensive overview and valuable insights into the field.

2. Literature Survey

2.1 Types of Modalities

Affective computing is an integrative field that traverses psychology, cognitive sciences, and computer science. For many years, sentiment analysis has become a vigorous domain of exploration in affective computing [1–3]. A company's success depends on customer satisfaction: if customers like the product, it's successful; if not, improvements are needed. To gauge success, analyzing customer opinions is crucial. This is where sentiment analysis comes in, defined as identifying opinions from various information sources. Both customers and manufacturers use sentiment analysis to understand product or service feedback, making it a key focus for businesses and researchers. However, text alone can miss nuances like sarcasm and humor. Incorporating multiple modalities, such as visual and acoustic signals, can more accurately determine sentiment polarity in these cases.

Multimodal sentiment analysis (MSA) involves analyzing sentiment using multiple data types, such as text, images, audio, and videos. In MSA, diverse data formats are processed by a model to predict sentiment. This research field is rapidly growing, driven by the proliferation of internet platforms like Facebook and Twitter, where people express opinions through various media. Given the vast amount of multimodal content, an efficient analysis system is essential. MSA can be bimodal, combining two modalities (e.g., text and images), or trimodal, combining three (text, audio, and images). This multimedia content offers richer information, making MSA a crucial area of study. Researchers now utilize text, images, and videos from social platforms to capture a more comprehensive understanding of sentiments.In light of these facts, the categorization of various modalities might be grouped, as shown in Fig. 1.

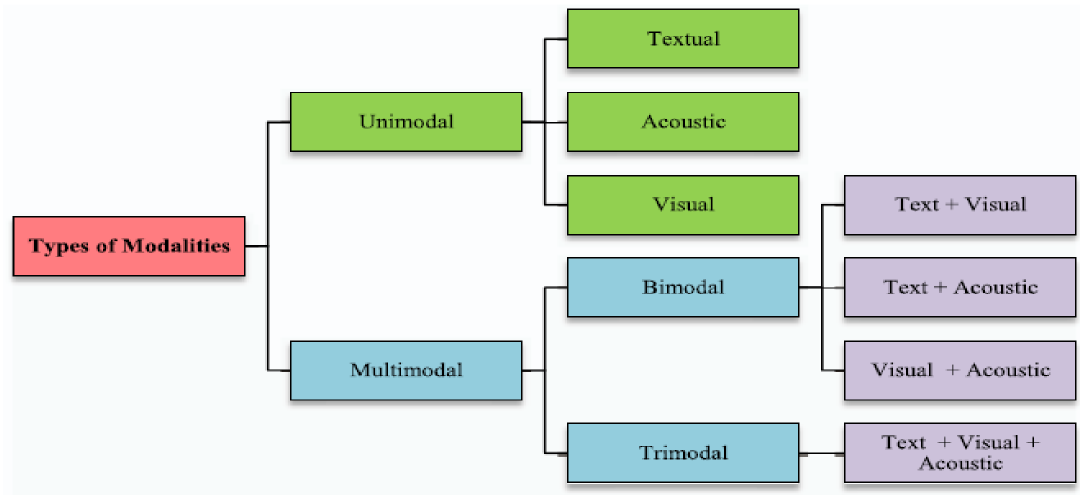

Fig 1: Types of Modalities used in MSA

The classification of data based on script-based information can be categorized into three distinct groups: document, sentence, and aspect- based. These three degrees of text-based information analysis are delineated in Fig. 2 with an example for better interpretation. Document-based classification involves distinguishing whether the entire assessment report is positive or negative. A product review, for example, is analyzed by the system to ascertain if its overall opinion is positive or negative. This sort of examination assumes that each record of the whole document expresses opinions about one particular unit (e. g., a unique merchandise or client service). Hence, it is not used for assessment reports that evaluate or compare multiple entities since such reports require a more detailed analysis similar to fine-grained ones. The cognitive science field is now applying various deep learning models to sentiment analysis at the document level, and recent work has been published in [4,5] to demonstrate this. On the other hand, sentence-based classification pertains to the determination of whether a given sentence is positive, negative, or neutral. There have been several publications on sentence level in recent years, among which the most recent one is [6]. The data in this study is first pre-processed to separate user comments from user ratings. On the basis of sentence-based SA, the comments of the users are categorized as positive, negative, or neutral. A Non-negative matrix factorization [7] technique has been utilized to obtain the positive, negative, and neutral topics for all the review comments available in the datasets. Similarly, topic distribution is done for the user rating data. And at last, with the help of a voting mechanism, sentiment prediction has been carried out.

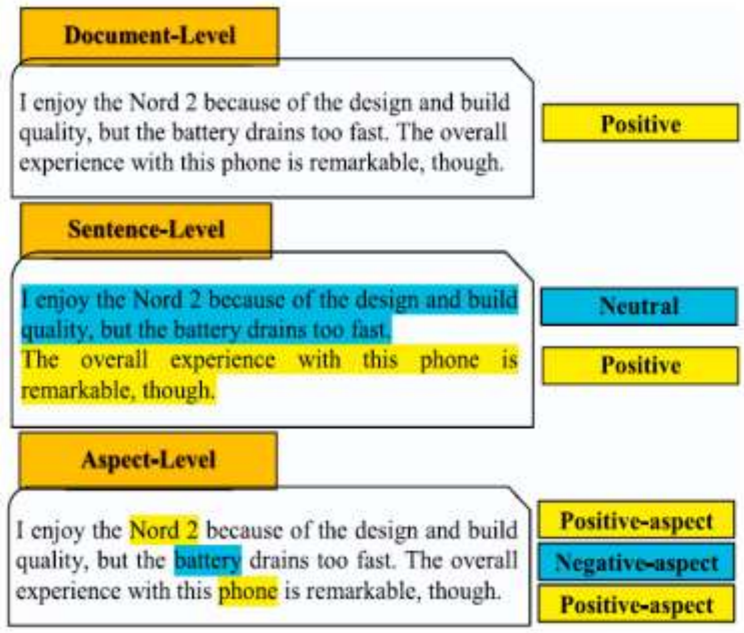

Fig 2: Three levels of text-based SA

Beyond the document and sentence levels, aspect-based analysis primarily centers on the identification of aspects, which refer to the attributes or characteristics present in a text-based review. This sort of sentiment analysis system is alluded to as fine-grained, which implies the detailed analysis of text-based information. It first attempts to extract the type of aspect and subsequently categorize the corresponding sentiments associated with that specific aspect. Deep learning models emerged as the ultimate winner out of various methodologies used for text-based sentiment analysis. Some exceptionally recent work incorporates [7–9]. Fig. 2 provides an example for distinguishing between the three levels of text-based SA.

The sentiment of the text, "I enjoy the Nord 2 because of the design and build quality, but the battery drains too fast. The overall experience with this phone is remarkable, though," is generally positive at the document level. At the sentence level, “I enjoy the Nord 2 because of the design and build quality, but the battery drains too fast” is neutral, while “The overall experience with this phone is remarkable, though” is positive. At the aspect level, the sentiment towards the Nord 2’s design and build quality is positive, but there is a negative concern about the battery life. Modern studies increasingly explore diverse datasets, including acoustic and visual data, as voice can provide richer emotional context than text alone. Acoustic sentiment analysis uses audio features to predict sentiment based on utterance characteristics. In [10], automatic sentiment analysis is done on naturalistic audio. Working on it is one of the most challenging and understudied research areas. Fig. 3 provides a visual representation of acoustic sentiment analysis, showcasing the process of obtaining mel-spectrograms and other features such as MFCC. These features are extracted using either the Librosa or OpenSmile toolkit. Subsequently, a deep ConvNet has been employed to extract pertinent acoustic features, which are then transmitted to the dense layer for the ultimate sentiment prediction. Research studies have not only directed their attention towards the identification of social anxiety but have also explored the potential of acoustic signals in detecting depression. In [11] depression is detected with the use of acoustic SA. This particular research work utilizes the DAIC-WOZ dataset.

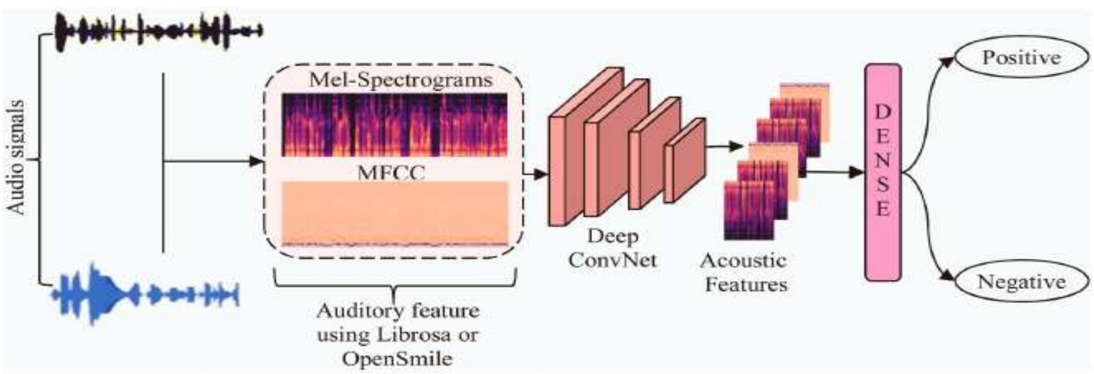

Fig 3: Acoustic sentiment analysis

In this study, first, the data is pre-processed to remove noise/silent features as well as the voice of the virtual interviewers from the audio, which are not useful. Then, the features from these pre-processed audios are extracted as spectrograms, and finally, all these spectrograms are fed into a 6-layer ConvNet for the actual prediction. [12] has utilized a multilingual dataset to perform sentiment analysis based on the audio. his study employs Short Time Fourier Transform to extract features from audio files, converting the resulting spectrograms into histograms using a bag-of-visual-words approach. The random forest ensemble learning method achieved the highest performance, with an accuracy of 76% and an F1 score of 78%. The growth of the World Wide Web has introduced numerous online platforms where users share multimedia content—audio, images, emojis, and videos—to express their sentiments. Recent research has expanded beyond text and audio to include visual data for sentiment analysis, allowing models to process images for emotion detection. The process of examining emotions and opinions through visual means is commonly known as visual sentiment analysis (Fig. 4). In the history of computer vision, deep learning models have proven their superiority for visual data [13], and all the research in visual sentiment analysis has primarily employed deep learning models [11–13]. Images are fed to deep learning models for visual sentiment analysis, which learn the visual features and score the sentiment polarity based on those features. The details of how features and models predict polarity are discussed in the following sections.

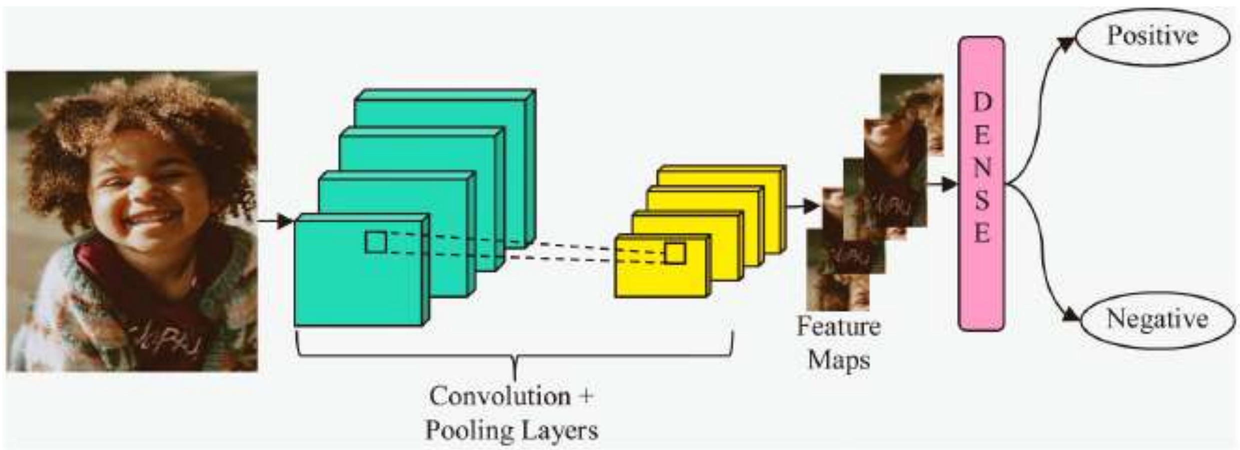

Fig 4: Visual sentiment analysis

In the current digital age, people encounter vast amounts of multimodal data on social media, including comments with images, text, audio, videos, and emojis. Extracting implicit emotions from such diverse data is challenging. While unimodal sentiment analysis can gauge explicit attitudes, it struggles with subtler emotions like sarcasm, humor, or exaggeration. To improve accuracy, multimodal data processing is preferred. This approach highlights the effectiveness of deep learning in addressing the complexities of multimodal sentiment analysis (MSA). The focus includes exploring various MSA applications, operations, deep learning architectures, fusion methods, challenges, benchmark datasets, and state-of-the-art techniques. [14–19] are a few recent publications that have established their relevance in this area. The model utilized in these research studies has integrated various modalities as input to estimate sentiment. A crisp idea of it is demonstrated in Fig. 5.

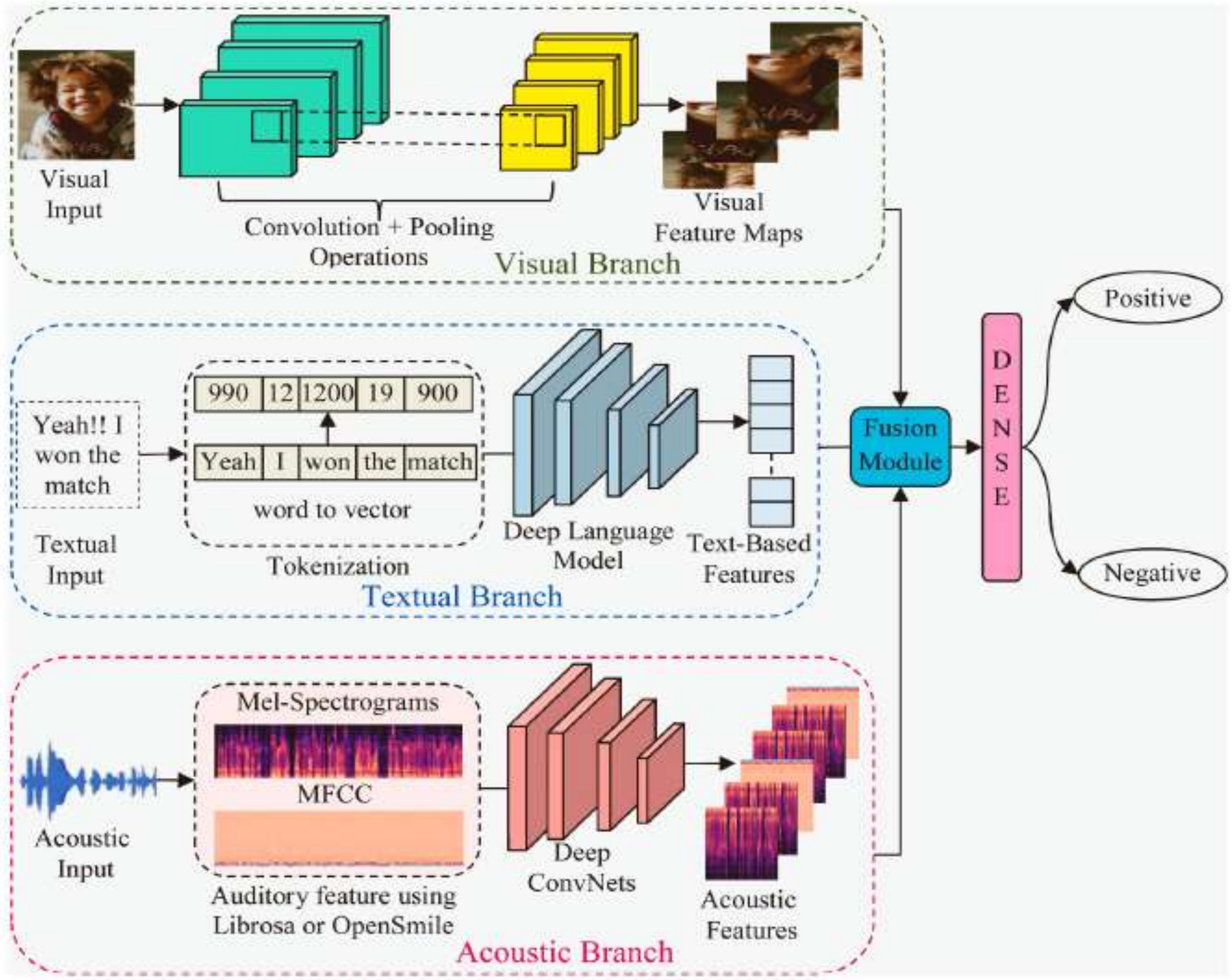

Fig 5: Multimodal sentiment analysis

2.2 Datasets

The following are the stages involved in creating a dataset for multimodal sentiment analysis.

Data Acquisition. The term ‘data acquisition’ refers to the act of gathering data from relevant sources before it is stored, cleansed, pre-processed, and used in other methods. For this purpose, videos are collected from several internet video sharing platforms for the creation of multimodal datasets. To crewel the movies from the web using specified search phrases, automatic or semi-automatic tools are utilized. To confirm that the video is a monologue, web videos are analysed for the presence of one speaker in the frame using facial detection. Typically, videos are chosen in which the speaker’s focus is solely on the camera. Videos are gathered based on frequently searched themes, and videos from each channel are limited to a certain quantity for greater variety.

Data Pre-Processing: Data often contains missing values, inaccuracies, or irrelevant properties, making pre-processing essential for quality improvement. For the MSA dataset, speakers are aged between mid-20s and late-50s, with a small number of fluent non-native English speakers and some wearing glasses. Few videos feature speakers with accents. Post-processing involves transcribing audio to integrate auditory, visual, and text data. Google Cloud Speech API and Amazon Transcribe provide suitable transcriptions, including nonverbal cues and contextual metadata like timestamps and punctuation, aiding in accurate alignment with annotations.

Data Annotation: Data annotation is the process of labelling images, video frames, audio, and text data that is mainly used in supervised machine learning and semi supervised machine learning to train the datasets. It helps machine to understand the input and act accordingly. For each clip, every annotator decides its sentimental state as -1 (negative), 0 (neutral) or 1 (positive). There are independent human resources who are used for making annotations. Then, in order to do both regression and multi-classifications tasks, average labelled result is used.

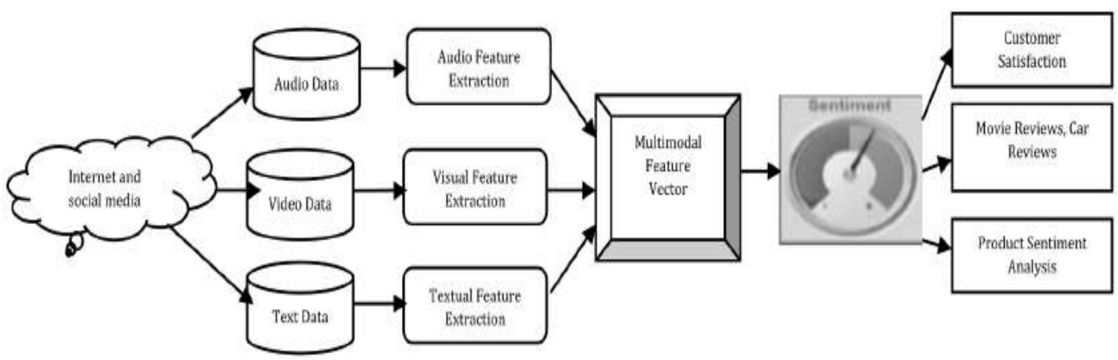

Fig 6: Fusion of multimodal data for sentiment analysis

Database	Year	Modalities	Source	Language	Topics	Available at
MOSI	2016	A+V+T	Youtube		#vlog	https://www.amir-zadeh.com/datasets
CMU-MOSEI	2018	A+V+T	Youtube	English	Review, debate,	https://www.amir-zadeh.com/datasets

					consulting	
MELD	2019	A+V+T	TV series, friends, reddit, facebook	English	Dialogues from TV series-Friends	https://affective-meld.github.io.
MULTI-ZOL	2019	V+T	ZOL.com	Chinese	Review of mobile phones	
CH-SIMS	2020	A+V+T		Chinese	General	
Memotion Analysis	2020	V+T	Reddit, Facebook , etc	English	52 Hillary Trump minion	*https://github.com/terenceylchow124/ Meme-MultiModal/tree/main /data/ memotion*
CMU-MOSEAS	2021	A+V+T	Youtube	Spanish, portugese, german French	General topics	https://github.com/thuiar/MMSA
MuSe-CaR	2021	A+V+T	Youtube	English	Vehicle review	*https://www.muse-challenge.org/challenge/ data*
B-T4SA	2021	V+T	Tweeter	Other than English	General	https://www.t4sa.it/
FACTIFY	2022	V+T	Tweeter	English	Politics, government	
MEMOTION 2	2022	V+T	Reddit, facebook	English	Politics, Religion, sports	

Table 1: Popular Datasets used for MSA

3.0 Research Methodology

The general process of a sentiment analysis system encompasses stages such as data collection, text preprocessing, feature extraction, training machine learning or deep learning models, and thorough evaluation to assess the model's effectiveness in sentiment classification tasks. Figure 7 explains a general process of prediction of sentiments.

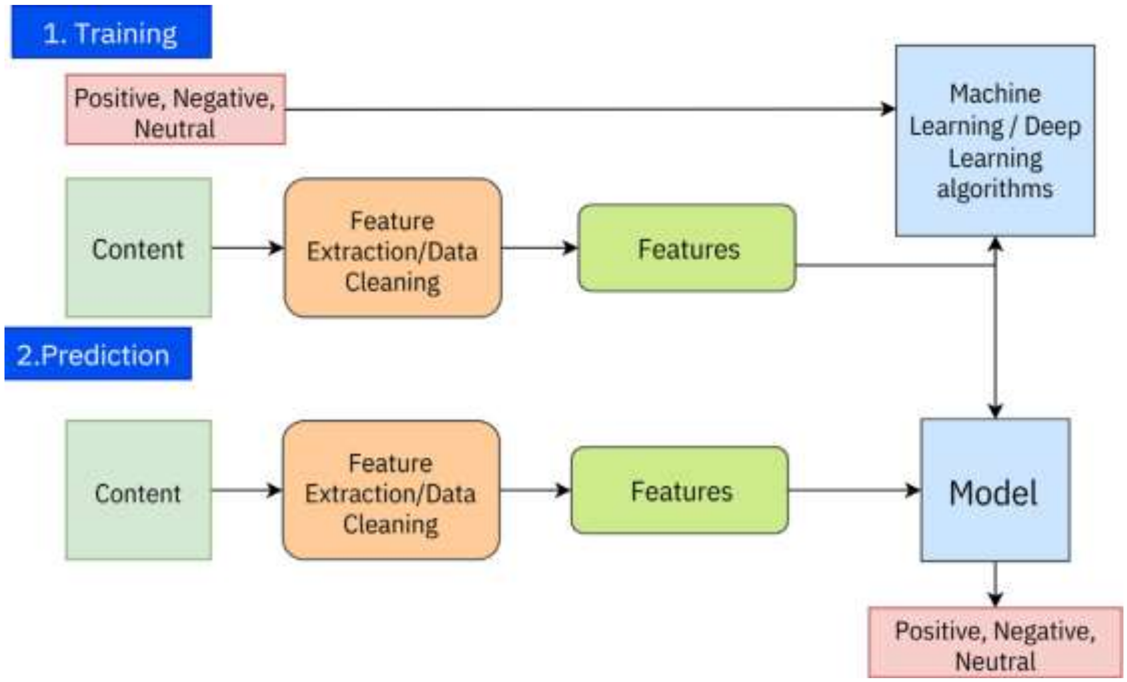

Fig 7: Generalized methodology for MSA

Various Machine Learning, Deep Learning and Large language Models are deployed for analyzing the sentiments. The focus will be on deep learning architectures, as presented in Fig. 8. Hence all the deep learning models that are used for MSA will be discussed here.

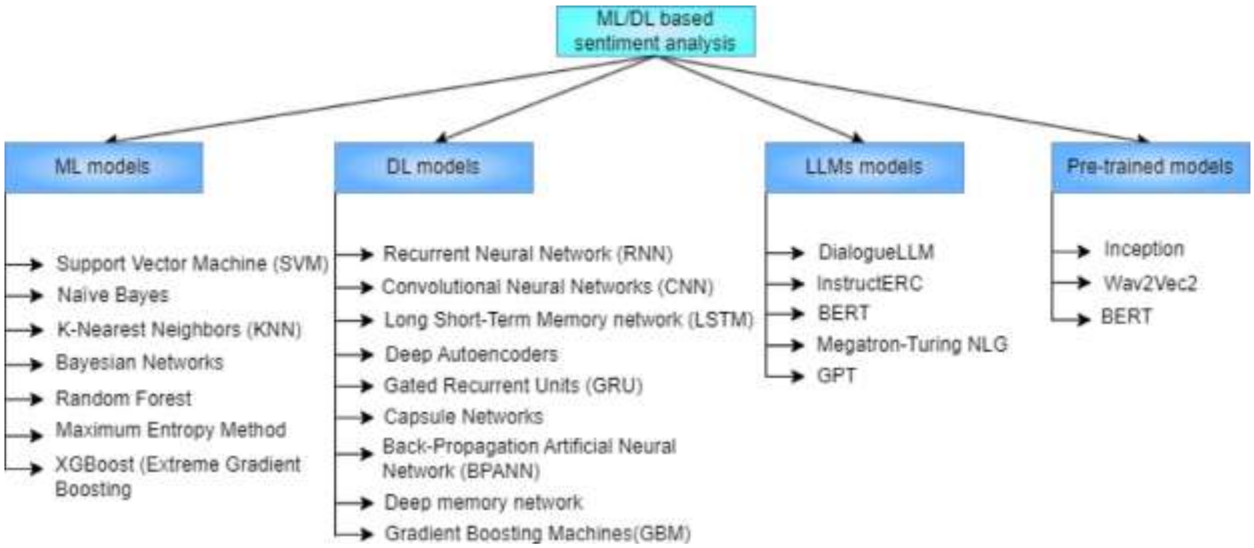

Fig: 8 MSA approached using machine Learning and Deep Learning algorithms

3.1 Critical Analysis of the Work done using various modalities and approaches

S. No.	**Title of the Paper**	**Author's Name**	**Year of Publishing**	**Objective (Problem Addressed)**	**Criteria /Source to Detect Depression**	**Data Set Used**	**Techniques Used to Detect Depression**	**Results / Parameters/ Discussions**
1	Identifying depression on Twitter	Nadeem, M.	2016	A method is established by which depression is recognized by analyzing large-scale records of users' linguistic histories on	Text-processing and Twitter Sentiment Analysis	2.5M Tweets	Decision Trees, Support Vector Classifier, Naive Bayes w/ 2-grams, Logistic	ROC AUC score of Logistic Regression 0.91, Linear SVM (0.80),

				social media, particularly Twitter			Regression, Naive Bayes w/ 1-grams and Ridge Classifier	Ridge Classifier (0.74), Decision Tree (0.64), Naive Bayes
2	Depression detection using emotion artificial intelligence	Deshpande, M., & Rao, V.	2017, December	NLP and sentiment analysis of tweets to detect depression.	Text-processing and Twitter Sentiment Analysis	10,000 Tweets	Multinomial Naïve Bayes algorithm, SVM	Precision, Recall, F1-Score, and Confusion Matrix
3	Predicting Depression Levels Using Social Media Posts	M.M. Aldarwish and H.F. Ahmad	2017	A system is suggested that leverages SNS as a data source and screening tool to classify users using AI based on user generated content on	Text-processing and Twitter Sentiment Analysis	Social Media posts (6773) from: LiveJournal, Twitter and Facebook	SVM and Naïve-Bayes	SVM Accuracy: 5, Precision: 6, Recall: 56%, Naive Bayes Accuracy: 6, Precisio

				SNS. SVM and Naïve-Bayes are the two different classifiers used to classify the UGC from SNS.				n: 100%, Recall: 58%
4	Detecting Depression in Social Media Posts Using Machine Learning	Biradar, A., & Totad, S.G.	2018, December	To create training data for the system, SentiStrength sentiment analysis has been done, and a BPNN model is utilized to classify the tweets into depressive or non-depressive categories.	Text-processing and Twitter Sentiment Analysis	BPNN and SentiStrength	A hybrid model that uses a sentiment analysis technique like SentiStrength to generate the train data and a BPNN model to categorize it.	Precision, Recall, F1-Score, and Confusion Matrix

5	Depression detection from social network data using machine learning techniques	Islam, Md Rafiqul, Kabir, Ashad, Ahmed, Ashir, Kamal, Abu, Wang, Hua, Ulhaq, Anwaar	2018	An investigation of depression is conducted using Facebook data obtained from an open public source. In Facebook users, machine learning techniques find high-quality solutions to mental health issues.	Text-processing	Facebook data (User's Comments)	Decision Tree, Support Vector Machine, Ensemble, and K-Nearest Neighbour	Precision, Recall, and measure (assessing emotional process, linguistic styles, Temporal process, and all features)
6	Detecting Depression Using K-Nearest Neighbors (KNN) Classification Technique	Islam, Md Rafiqul, Kamal, Abu, Sultana, Naznin, Islam,	2018	Detecting Facebook users' moods using the KNN (k-nearest neighbours) classification algorithm. Fine KNN,	Text-processing	Facebook data	Fine KNN, Medium KNN, Coarse KNN, Cosine KNN, Cubic KNN,	Precision, Recall, and measure (assessing emotional

		Robiul, Ulhaq, Anwaar , Moni, Moham mad		Medium KNN, Coarse KNN, Cosine KNN, Cubic KNN, and Weighted KNN were employed as KNN Classifiers.			and Weighte d KNN	process , linguist ic styles, Tempor al process , and all features)
7	Depression Detection With Sentiment Analysis Of Tweets	Hemant hkumar M, Latha A.	2019, May	Sentiment analysis of Twitter feeds and classification as positive, neutral, and negative. Dataset with 43,000 tweets is taken and tweets are divided into ratio 70:30.	Text-processi ng and Twitter Sentime nt Analysis		Multino mial Naïve Bayes algorith m, SVM	Precisio n, Recall, and measur e (assessi ng emotio nal process , linguist ic styles, Tempor al process , and all features

								)
8	Study of Depression Analysis using Machine Learning Techniques	Devakunchari Ramalingam, Vaibhav Sharma, Priyanka Zar.	2019, May	Various machine learning techniques used to detect depression analyzing tweets and various social media posts. In-time perception in depression, filtering, and classification processes are discussed.	Text-processing		Various ML Techniques	Precision, Recall, and measure (assessing emotional process, linguistic styles, Temporal process, and all features)

9	Depression Detection of Tweets and A Comparative Test	Rajaraman, Nath, Akshaya.P.R, Bhuja. G	2020, March	Comparative analysis of various techniques like TF-IDF, Naïve-Bayes, LSTM, Logistic Regression, Linear Support Vector and depression through tweets is being detected.	Text-processing and Twitter Sentiment Analysis	Sentiment140, tweet Scrap from TWINT and Google Word2Vec	TF-IDF, Naïve-Bayes, LSTM, Logistic Regression, Linear Support Vector, Deep learning model using NLP	Precision, Recall, F1-score, Support, and Accuracy
10	Machine Learning-based Approach for Depression Detection in Twitter Using Content and Activity Features	AlSagri, H.S., & Ykhlef, M.	2020	Various ML-based approaches are exploited to find whether a Twitter user is depressed or not based on his/her social network behavior and tweets.	Text-processing and Twitter Sentiment Analysis	Tweets	SVM, Naïve-Bayes, and Decision Trees	Precision, Recall, F1-measure, and Area Under Curve (AUC)

				‘‘Depression detection using activity and content features (DDACF)” classification model is used.				
11	Predicting depression using deep learning and ensemble algorithms on raw Twitter data	Shetty, N.P., Muniyal, B., Anand, A., Kumar, S., & Prabhu, S.	2020	ML classifiers are employed in the Twitter dataset to identify whether a person is depressed or not.	Text-processing and Twitter Sentiment Analysis	Twitter Data set	LSTM, CNN, linear support vector classifier (SVC), multinomial naïve-bayes, Bernoulli naïve-bayes, logistic regression, random forest classifier, and	Precision, Recall, F1-measure

							gradient boosting classifier.	
12	Face-Based Automatic Human Emotion Recognition	Savadi, VPatil	2014, July	YALE Database is considered. Image processing and emotion recognition through facial expression.	Facial Expressions (Image and Video Processing)	A video of 810 frames	Neural Networks, Logarithmic Gabor filters, FACS, skin mapping, pattern matching, etc.	Precision, Recall, F1-measure, and Area Under Curve (AUC)
13	Image Processing Techniques To Recognize Facial Emotions	A. Mercy Rani, R. Durgadevi	2017, August	Emotion recognition through facial expression. Recognizing facial emotions based on filled mouth-regions.	Facial Expressions (Image and Video Processing)	Facial Expressions (Image and Video Processing)	Neural networks, Viola-Jones algorithm, etc.	Precision, Recall, F1-measure

				Viola-Jones algorithm is used.				
14	Comparative Study on Mood Detection Techniques	Gavde, Megna	2018	Comparative study of different methods to recognize emotion from facial expressions.	Facial Expressions (Image and Video Processing)	PCA, Gabor Filters, SVM, K-means clustering, etc.	Gabor feature extraction is used to extract features from face. Trained neural network is used for face detection. SVM is used for emotions classification.	Accuracy: Effective (quantitative value given)

15	Depression Detection and Analysis	Oak, S.	2017, March	Detection of depression in an individual through RBFN. Text and speech input are considered. Partial Least Square algorithm is used to detect depression from vocal expressions.	Use Of Chatbots, Emotional AI and Combined Inputs (Text, Audio, Image, Video)	53 Volunteers	RBFN (Radial basis function networks)	Accuracy: 71.4%
16	Review on Mood Detection using Image Processing and Chatbot using Artificial Intelligence	Thosar, Gothe, Bhorkade, Sanap	2018, March	Mood detection through image processing using Haar-cascade algorithm, chatbots to detect stress through text/speech input.	Use Of Chatbots, Emotional AI and Combined Inputs (Text, Audio, Image, Video)	An image with 6000 features	Haar-cascade algorithm and Ad-boost	Precision: Satisfactory

17	Detecting depression using a framework combining deep multimodal neural networks with a purpose-built automated evaluation	Victor, E., Aghajan, Z.M., Sewart, A.R., & Christian, R.	2019	Artificial Intelligence based Mental Evaluation system can predict whether the patient is depressed or not through deep learning.	Use Of Chatbots, Emotional AI and Combined Inputs (Text, Audio, Image, Video)	671 participants recorded by webcam and microphone while responding interview questions	Multimodal Deep Learning Model	Precision: 68.61, NPV: 67.95, Sensitivity: 68.59, Specificity: 67.46, Score: 67.6
18	The utility of artificial intelligence in suicide risk prediction and the management of suicidal behaviors	Fonseka, T.M., Bhat, V., & Kennedy, S.H.	2019	This study includes an overview of the literature as well as the role of AI in predicting and managing suicide risk.	Use Of Chatbots, Emotional AI and Combined Inputs (Text, Audio, Image, Video)	ML methods, AI and conversational agents	Predictive models in ML can help to detect suicide risk. AI has been linked with social media to detect	Precision: Satisfactory

							and analyse depression in social media users.	
19	AI Therapist Using Natural Language Processing	Shephali Santosh Nikam, Aishwarya Vijay Patil, Gauri Shashikant Patil, Sharvari Pramod Patil, B. D. Jitkar.	2020, February	An Expert System for Stress Management to provide best recommendation and solution for youth, especially IT professionals. NLP, Naïve-Bayes, collaborative-filtering algorithms are used. Back-propagation method is used to train the system.	Use Of Chatbots, Emotional AI and Combined Inputs (Text, Audio, Image, Video)	Chatterbot Corpus	Naïve-Bayes Algorithm and Collaborative filtering algorithm, chatbots and Back-propagation method	Precision: 68.61, NPV: 67.95, Sensitivity: 68.59, Specificity: 67.46, Score: 67.6

20	The Utility of Artificial Intelligence for Mood Analysis, Depression Detection, and Suicide Risk Management	Zohuri, Bahman & Zadeh, Siamak.	2020	Detection of depression by various techniques: combining OCR with AI, deep learning models.	Use Of Chatbots, Emotional AI and Combined Inputs (Text, Audio, Image, Video)	Combined application of Image Processing, Voice or speech recognition and AI, deep learning models	Combining Optical Character Reader with AI leads to facial emotion recognition and can detect whether a person in an image or video is smiling or sleeping (eyes closed) and this can lead to early detection of depressi	Precision: 68.61, NPV: 67.95, Sensitivity: 68.59, Specificity: 67.46, Score: 67.6

							on and preventi on of suicide.	
21	Mentalbert: Publicly available pretrained language models for mental healthcare.	Shaoxi ong Ji, Tian lin Zhang, Luna Ansari, Jie Fu, Pra yag Tiwari, Erik Cambri a	2021	This paper trains and release two pretrained masked language models, i.e., MentalBERT and MentalRoBE RTa, to benefit machine learning for the mental healthcare research community.	Tokeniz ation, Removi ng, Data Cleanin g	Reddit, Twitter	Mental BERT, Mental RoBER Ta	Mental RoBER Ta: F1 Score: 94.62% , Mental BERT: F1 Score: 81.76%

22	Crass: A novel data set and benchmark to test counterfactual reasoning of large language models.	Frohberg, J., Binder, F.	2021	We introduce the CRASS (counterfactual reasoning assessment) data set and benchmark utilizing questionized counterfactual conditionals as a novel and powerful tool to evaluate large language models.	Features extraction	Social Media, Ecommerce, CRASS	MPNet, GPT-3	Accuracy: GPT-3: 31.85% , MPNet: 35.11%
23	A comprehensive survey on sentiment analysis: Approaches , challenges and trends	Birjali et al.	2021	presents a complete study of sentiment analysis approaches, challenges, and trends, to give researchers a global survey on sentiment analysis and its related	Removing, Repeated , Characters, Features Extraction, Features Selection	Social media	SVM, ANN	Accuracy: 92.17%

				fields.				
24	ANTi-Vax: a novel Twitter dataset for COVID-19 vaccine misinformation detection	K. Hayawi a, S. Shahriar a, M. A. Serhani b, I. Taleb a, S.S. Mathew a	2022	The goal of this research is to introduce a novel machine learning–based COVID-19 vaccine misinformation detection framework.	Stemming	ANTi-Vax	BERT	Accuracy: 98%, F1-score: 98%, Precision: 97%,Recall: 98%
25	TSA-CNN-AOA: Twitter sentiment analysis using CNN optimized via arithmetic	Aslan, S., Kızıloluk, S., Sert, E.	2023		Removing, Tokenization, Lemmatization, Stemming	Twitter API	Logistic regression, Decision tree	Accuracy: LR: 88.94%, DT: 92.53% F1-score:

	optimization algorithm							LR: 85.71%, DT: 90.54%
26	COVID-19 vaccine hesitancy: Text mining, sentiment analysis and machine learning on COVID-19 vaccination Twitter dataset.	Qorib, M., Oladunni, T., Denis, M., Ososanya, E., Cotae, P.	2023	Thus, this study examines 3 sentiment computation methods (Azure Machine Learning, VADER, and TextBlob) to analyze COVID-19 vaccine hesitancy.	Tokenization, Removing, Stemming, Lemmatization	Twitter COVID-19	TF-IDF +, LinearSVC, Random Forest, Logistics Regression, Decision Tree, LinearSVC, and Naïve Bayes)	Accuracy: 96.75% ,Precision: 96.92%
27	Using sentiment analysis to evaluate qualitative students' responses.	Dake, D.K., Gyimah, E.,	2023	sentiment analysis helps understand learners' appreciation of lessons, which they prefer to	Tokenization, Data removing	Qualitative Feedback from students	Naïve Bayes (NB), Support Vector Machine (SVM),	Accuracy: 63.79%

				express in long texts with little or no restriction. Such expressions depict the learner's emotions and mood during class engagements.			J48 Decision Tree (DT), and Random Forest (RF)	
28	LLM augmented LLMs: Expanding capabilities through composition.	Bansal, R., Samanta, B., Dalmia, S., Gupta, N., Vashishth, S., Ganapathy, S., Bapna, A., Jain, P., Talukdar, P	2024	we propose CALM -- Composition to Augment Language Models -- which introduces cross-attention between models to compose their representations and enable new capabilities.	Tokenization, Feature Extraction	GSM-8K	CALM Composition to Augment Language Models	Accuracy: 84.3%

4. Challenges

Ensuring user privacy and obtaining informed consent are significant challenges. Researchers must navigate the ethical implications of using personal data for mental health analysis (Moreno et al., 2013). Models trained on specific datasets may not generalize well across different populations or platforms. Ensuring robustness and accuracy in diverse contexts remains a challenge (Calvo et al., 2017). Integrating text, image, and user interaction data can provide a more comprehensive view of users' mental health (Yang et al., 2018). Developing systems for real-time monitoring and intervention can help in providing timely support to individuals at risk (Conway & O'Connor, 2016). Conducting studies across different cultural contexts can help in understanding the universal and culture-specific markers of depression (Coppersmith et al., 2014) Depression is a dynamic condition, and social media behavior can change over time. Capturing these temporal dynamics is crucial for accurate detection (Schwartz et al., 2014).

5. Applications

Sentiment analysis has diverse applications across industries. It offers valuable insights into public sentiment and opinions. Critical applications include customer feedback analysis for improving products and services, real-time social media monitoring for brands and individuals, market research to assess consumer trends, brand reputation management to prevent crises, political campaigning for understanding public opinion, and aiding product development. It is also used in financial markets, healthcare, and media outlets to gauge content impact and academic research to study public attitudes and social trends. Sentiment analysis is vital in understanding and responding to public assessing sentiment in the contemporary data-driven landscape across diverse domains. Figure 9 shows various applications of sentiment analysis across myriad domains.

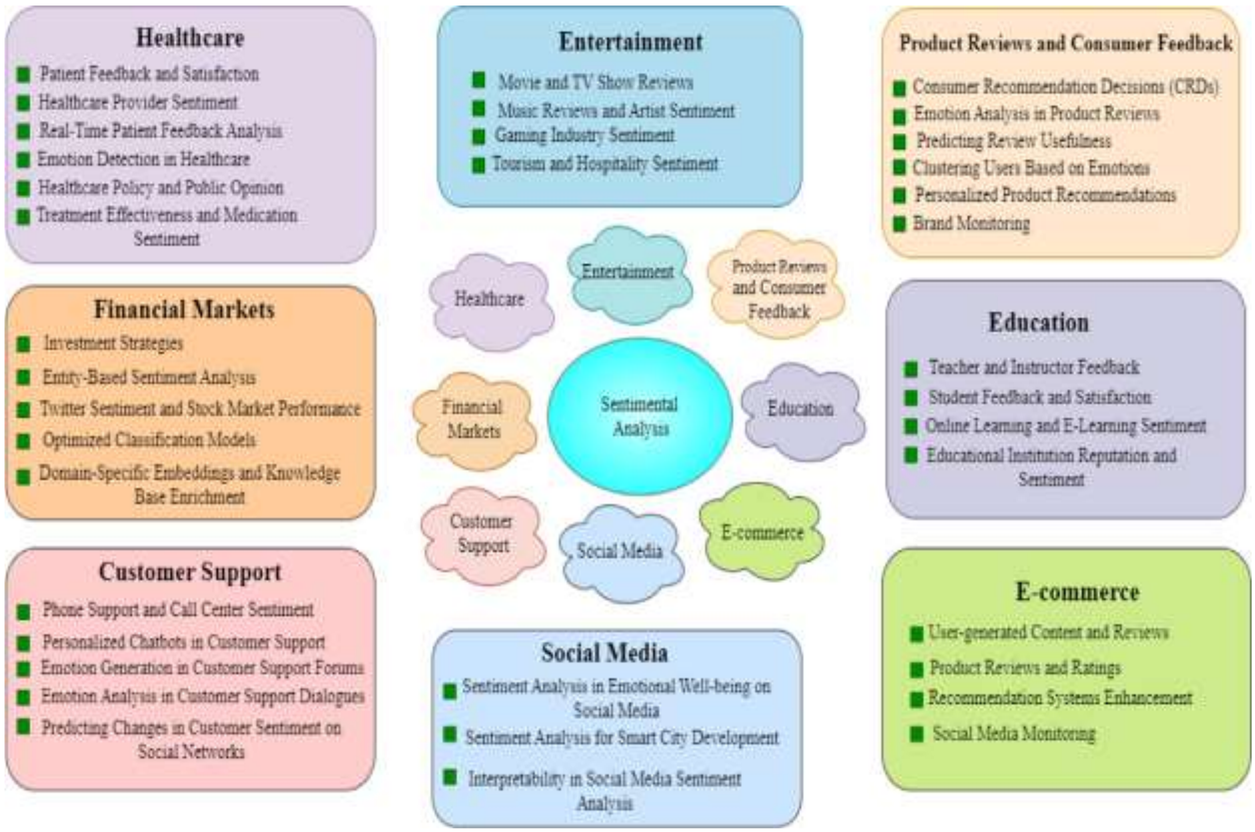

Fig: 9 Applications of sentiment analysis

6. Conclusion

Depression detection from social media is a rapidly evolving field with significant potential for early intervention and support. While there are challenges to be addressed, advancements in NLP, machine learning, and multimodal analysis hold promise for improving the accuracy and applicability of these methods. Ethical considerations and privacy concerns must remain at the forefront to ensure responsible use of social media data in mental health research.

An automatic technique for recognizing the polarity of opinions is called sentiment analysis. Multimodality is defined as the analysis of more than one mode of data. The term MSA refers to the incorporation of more than one input mode in order to improve the accuracy of the analysis; tens of thousands of videos are uploaded online every day, and analysis of all these media is increasingly important; analyzing multimodal opinions automatically requires a profound understanding of natural languages, audio, and video processing, whereas researchers are constantly working to improve them. This article covers the most important work on MSA using deep learning architectures. To begin with, we discussed multiple types of modalities involved in MSA, followed by their applications or significance. We gathered data for the review from the most well-known research publications from 2017 to 2023. Additionally, this survey covers different deep-learning models, fusion strategies, and operations associated with MSA. Our research has concluded that the Transformer-based models are the clear winners in this field, as they have provided the best results till date. Furthermore, a list of almost all the publicly available video datasets and image-text pair datasets for MSA has been provided. Along with the whole comprehensive list of datasets, we also select the best cutting-edge models tested so far along with the benchmark datasets. These days, people prefer to use multimodal data to gather sentiments; hence, MSA holds great promise for future applications such as E-learning, emoji prediction, E-governance, temporal tagging, etc.

References

[1] L. Teijeiro-Mosquera et al., "What your face vlogs about: Expressions of emotion and big-five traits impressions in YouTube," '*IEEE Trans. Affective Comput'* vol. 6, no. 2, pp. 193-205, 2015.

[2] E. Cambria, "Affective computing and sentiment analysis, IEEE Xplore,"*J. Intell. Syst.*,vol.31,no. 2, pp. 102-107, 2016.

[3] E. Cambria et al., "Sentiment analysis is a big suitcase, IEEE Xplore," J. Intell. Syst., vol. 32, no. 6, pp. 74-80, 2017.

[4] S. Liu and I. Lee, "Sequence encoding incorporated CNN model for Email document sentiment classification,''*Appl. SoftComput.*,vol.102, p. 107104,2021

[5] Y. Zhang et al., "Conciseness is better: Recurrent attention LSTM model for document-level sentiment analysis,'' *Neurocomputing*, vol. 462, pp. 101-112, 2021.

[6] C. Yang et al., "Leveraging semantic features for recommendation: Sentence-level emotion analysis,'*Inf.Process.Manag.*, vol.58, no. 3, p. 102543,2021.

[7] J. Wang et al., "Research hotspot prediction and regular evolutionary pattern identification based on NSFC grants using NMF and semantic retrieval," *IEEE Access*, vol. 7, 123776-123787, 2019.

[8] B. Liang et al., "Aspect-based sentiment analysis via affective knowledge enhanced graph convolution networks," *Knowl. Based Syst.*, vol. 235, p. 107643, 2022.

[9] H. Wu et al., "Phrase dependency relational graph attention network for aspect-based sentiment analysis, "*Knowl. Based Syst.*, vol. 236, p. 107736, 2022.

[10] L. Kaushik et al., "Automatic sentiment detection in naturalistic audio" I*EEE/ACM Trans. Audio Speech Lang. Process* ,vol. 25, no. 8, pp.1668-1679, 2017.

[11] H. Negi et al., "A novel approach for depression detection using audio sentiment analysis," Int. J. Inf. Syst. Manag. Sci., vol. 5068, p. 1556, 2018.

[12] S. Luitel and M. Anwar, "Audio sentiment analysis using spectrogram and bag of words" in *IEEE 23rd International Conference on Information Reuse and Integration for Data Science (IRI)* 2022, pp. 6655-6603.

[13] J. Joo et al., "Visual persuasion: Inferring communicative intents of images", in *Proc. 2014 IEEE Conference on Computer Vision and Pattern Recognition, IEEE*, Jun. 2014, pp. 216-223.

[14] A. Pandey and D.K. Vishwakarma, VABDC-net: A framework for visual-caption sentiment recognition via spatio-depth visual attention and bi-directional caption processing, Knowl. Base based Syst. (2023) 110515.

[15] A. Pandey and D.K. Vishwakarma, "Attention-based model for multi-modal sentiment recognition using text-image pairs" in *Proc. 2023 4th International Conference on Innovative Trends in Information Technology (ICITIIT)*.

[16] J. Xu, et al., "Visual-textual sentiment classification with bi-directional multi-level attention networks, "*Knowl. Based Syst.*, vol.178, pp. 61-73, 2019.

[17] J. Yu et al., "Entity-sensitive attention and fusion network for entity level multimodal sentiment classification, "*IEEE/ACM Trans. Aud. Speech Lang. Process,* vol.28, pp. 429-439, 2020.

[18] S. Mai et al., “Analyzing multimodal sentiment via acoustic- and visual- LSTM with channel-aware temporal convolution network, ”*IEEE/ACM Trans. Audio Speech Lang. Process,* vol. 29, pp. 1424-1437, 2021.

[19] V. Lopes et al., “An Auto ML-based approach to multimodal image sentiment analysis” in *2021 International Joint Conference on Neural Networks (IJCNN)*, IEEE, 2021.

[20] D. Gu, et al., “Targeted aspect-based multimodal sentiment analysis: An attention capsule extraction and multi-head fusion network,”*IEEE Access*, vol. 9, 157329-157336, 2021.

[21] A. Pandey, D.K. Vishwakarma, VABDC-net: a framework for visual-caption sentiment recognition via spatio-depth visual attention and bi-directional caption processing, Knowl. Based Syst. 2023.

[22] A. Pandey and D. K. Vishwakarma, “Attention-based model for multi-modal sentiment recognition using text-image pairs” in Proc. 2023 4th International Conference on Innovative Trends in Information Technology (ICITIIT), IEEE, 2023, pp. 1-5.

[23] S. A. Abdu et al., “Multimodal Video Sentiment Analysis Using Deep Learning approaches, a Survey.” Inf. Fusion, vol. 76, pp. 204-226, ISSN 1566-2535, 2021.

[24] A. Yadav and D. K. Vishwakarma, “Sentiment analysis using deep learning architectures: A review” Artif. Intell. Rev., vol. 53, no. 6, pp. 4335-4385, 2020.

[25] Calvo, R. A., Milne, D. N., Hussain, M. S., & Christensen, H. (2017). Natural language processing in mental health applications using non-clinical texts. *Natural Language Engineering, 23*(5), 649-685.

[26] Chancellor, S., Birnbaum, M. L., Caine, E. D., Silenzio, V. M. B., & De Choudhury, M. (2019). A taxonomy of ethical tensions in inferring mental health states from social media. *Proceedings of the ACM on Human-Computer Interaction, 3*(CSCW), 1-30.

[27] Conway, M., & O'Connor, D. (2016). Social media, big data, and mental health: current advances and ethical implications. *Current Opinion in Psychology, 9*, 77-82.

[28] Coppersmith, G., Dredze, M., & Harman, C. (2014). Quantifying mental health signals in Twitter. *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, 51-60.

[29] De Choudhury, M., Counts, S., & Horvitz, E. (2013). Predicting postpartum changes in emotion and behavior via social media. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 3267-3276.

[30] Guntuku, S. C., Yaden, D. B., Kern, M. L., Ungar, L. H., & Eichstaedt, J. C. (2017). Detecting depression and mental illness on social media: an integrative review. *Current Opinion in Behavioral Sciences*, *18*, 43-49.

[31] Moreno, M. A., Goniu, N., Moreno, P. S., & Diekema, D. (2013). Ethics of social media research: common concerns and practical considerations. *Cyberpsychology, Behavior, and Social Networking*, *16*(9), 708-713.

[32] Reece, A. G., & Danforth, C. M. (2017). Instagram photos reveal predictive markers of depression. *EPJ Data Science, 6*(1), 15.

[33] Reece, A. G., Reagan, A. J., Lix, K. L. M., Dodds, P. S., Danforth, C. M., & Langer, E. J. (2017). Forecasting the onset and course of mental illness with Twitter data. *Scientific Reports, 7*(1), 13006.

[34] Resnik, P., Armstrong, W., Claudino, L., Nguyen, T., Nguyen, V. A., & Boyd-Graber, J. (2015). Beyond LDA: exploring supervised topic modeling for depression-related language in Twitter. *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, 99-107.

[35] Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M. & Ungar, L. H. (2014). Toward assessing changes in degree of depression through Facebook. *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, 118-125.

[36] Shen, G., Jia, J., Nie, L., Feng, F., Zhang, C., Hu, T., ... & Zhu, T. (2017). Depression detection via harvesting social media: A multimodal dictionary learning solution. *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, 3838-3844.

[37] Tsugawa, S., Kikuchi, Y., Itoh, Y., Fushimi, M., & Ohsaki, H. (2015). Recognizing depression from Twitter activity. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 3187-3196.

[38] Yang, S., Huang, T., Shen, W., Zhou, G., & Li, J. (2018). Automatic detection of depression and anxiety through digital indicators: A systematic review. *Proceedings of the 9th International Conference on Information and Communication Technologies for Health, Education, and Social Services*, 1-7.