TEXTUAL-FACIAL EMOTION DETECTION USING GROUP DECISION CLASSIFICATION AND DECISION LEVEL FUSION

Vandana Devi [1], Bhupesh Gupta [2] and Avinash Sharma [3]

[1] PhD Scholar, Computer Science & Engineering, Maharishi Markandeshwar Deemed University Mullana, Ambala, Haryana, India.
[2,3] Professor, Computer Science & Engineering, Maharishi Markandeshwar Deemed University Mullana, Ambala, Haryana, India.
Email: [1]vandukhanchi@gmail.com, [2]bhupeshgupta81@gmail.com, [3]asharma@mmumullana.org

DOI: 10.5281/zenodo.10964469

Abstract

In today's world the use of social networking sites is increasing consistently and a huge amount of data has been generated due to online community communication. This research has applications in boosting customer satisfaction e learning psychological Healthcare and building smartphones and other gadgets that can sense human emotions. Emotion recognition is a very tough research subject due to its complexity, as individual variances in cognitive-emotional cues entail a wide variety of methods, including language, expressions, and voice. A multimodal approach augmented by sentiment detection technology has the potential to revolutionize human-machine interaction. The fusion of text and face data allows for a more comprehensive understanding of human communication and behavior. By combining textual information such as written words with facial expressions and visual cues, the system can capture a richer set of features and context for analysis. In response to the above problems, this paper proposes a framework for a multimodal emotion detection method that includes feature extraction from both modalities text and face, followed by an ensemble-based group decision classifier method and last decision level fusion occurs which combines classification outputs from each modality. This paper's proposed method for text and face fine-grained sentiment analysis outperforms individual classifiers in terms of accuracy (97%) and F1-measure (94%) value.

Keywords: Emotion, Machine Learning Models, Text Preprocessing, VGG 16.

I. INTRODUCTION

People utilize the internet more often because of online media and commerce, especially given how widely cell phones are used. These encourage businesses to learn from people who are located far away. As a result, the amount of information on the Internet has grown dramatically. It considers emotions before making a choice. On social media platforms, users discuss their beliefs, concepts, attitudes, sentiments, and more in well-known languages. Thanks to the rapid growth of web-based media platforms like Facebook, Twitter, Quora, and others, anyone can publish their ideas online for anyone else to exploit for their own ends [1]. Twitter's significance as an arena for sentiment analysis research is owing to the amount of users and the accessibility of text data with opinions. Despite being brief and lacking in a consistent grammatical framework, these writings often reveal insightful user emotional inclinations [2]. Besides marketing and psychology, there is a tremendous amount of potential use for emotion detection in text. Commercial institutions, political campaigns, and organizations tasked with directing the reaction to a natural disaster can all profit from this. Additionally, AI systems, such as chatbots, can benefit from a grasp of emotions. [3]. Facial expressions are the most convincing form of nonverbal communication since they reveal one's intentions, level of understanding, and emotional state. Listeners can send a wealth of information to the speaker without even having to utter a word thanks to the power of facial expressions, which can shift the path of a conversation. [4]. Emotions may be more vividly and dynamically represented in multimodal information than in single-modal content. Cutting-edge research in multi-modal emotion detection lies at the crossroads of AI, computer vision, and NLP. It compiles and analyzes many forms of data to learn about people's emotions [5]. Annotated datasets with labels of emotions, text-image interactions, and emotion triggers are necessary for research into the functions of images in social media posts [6]. Word embedding in neural networks allows deep learning networks to learn accurate vector representations for sentiment analysis and emotion detection. When there aren't enough resources available or there's a shift in the domain, emotion detection can benefit from transfer learning [7]. The use of virtual reality (VR) in emotion recognition research is growing because it enables the replication of scenarios under controlled laboratory conditions with a strong sense of presence and engagement [8]. Sentence-level, document-level, and aspect-level sentiment analysis is divided into three tiers. Sentences inside texts or paragraphs are broken down, and the polarity of each sentence is determined at the sentence or phrase level of sentiment analysis. The sentiment is identified across the board in the document or record at the document level. Sentiment analysis assesses opinions regarding particular aspects or features at the aspect level [9]. The goal is to leverage the strengths of each individual model and create a more accurate and robust prediction. Different approaches and techniques can be used to combine the outputs of the individual models, such as voting, averaging, or weighted averaging. The specific method of combination depends on the problem at hand and the characteristics of the individual models [10].

Sentiment analysis currently employs a wide range of methods, such as models, constrained local models, semantic and rule-based approaches, dictionary-based lexicon approaches, corpus-based lexicon approaches, incremental naive Bayes classifiers, machine-learning-based filters, bag-of-words approaches, case conversion, N-gram removal, stemming, and segmentation [11].

Several data fusion methods are used in multimodal emotional identification. Some examples are fusion based on deep learning, fusion based on attention, fusion based on hybrid multimodal data, fusion based on features, and fusion based on decisions [12].

This research article's main objective is to identify various emotions based on text and face traits by utilizing textual and facial data. Four distinct emotional expression types are used to achieve this. Multimodal data, group decision categorization, and decision-level fusion techniques all offer better accuracy than the existing method when compared to it.

The structure of the paper is organized in five sections where section 1. deals with the introduction to the paper. Section 2. describes the related work in the related field. Section 3. Explain the proposed framework. Section 4. describes and analyses the experimental results. Section 5. conclusion and future Work.

The main contributions of this work are as follows.

- Data collection and preprocessing
- Feature extraction
- To apply different machine learning algorithms on our datasets and group decision classification and after that decision-level fusion.

II. RELATED WORK

This literature review aims to investigate the current state of the art in multi-modal emotion detection using text and facial images. This review will focus on the methods, techniques, and algorithms used in previous research to combine these modalities and extract meaningful emotional insights. This review seeks to identify trends, challenges, and advancements in the field by analyzing the results of various research projects. This review examines a variety of studies that investigate the integration of text and facial image data for emotion recognition. We will pay particular attention to studies that employ classifiers such as Support Vector Machine (SVM), Decision Tree (DT), K-Nearest Neighbours (KNN), and Ensemble as well as deep learning architectures such as VGG. Additionally, a comparative analysis of reported accuracy or F1-Measure will be conducted to determine the efficacy of various methodologies and models. This work focuses on sentiment analysis with deep learning, which is a novel approach in comparison to other methods. The study suggests a CNN-TDIDF family model, which is built on deep learning and integrates CNN and TF-IDF in tandem. Tests conducted on multiple challenging datasets reveal that the recommended methodology outperforms a lot of common methods. It has been demonstrated that this model can get an 87% accuracy rate and performs better than traditional machine learning techniques. In addition, the study achieved a very high accuracy rate in comparison to earlier attempts [13]. This study looks into what happens when you change the expressions on your profile pictures and how that affects text-based contact in fully remote workplaces. The researchers looked into how people's facial emotions affected how they understood neutral, positive, and negative messages. The study found that the expression on a profile picture's face changes how people feel about the sender and how they interpret neutral and positive messages. [14]. In order to fulfill this requirement, this article introduces a framework that provides a foundation for conducting a comparative assessment of fusion methods. The framework illustrates how these methods adjust to variations in quality among individual modalities and assess their overall performance. The findings indicate that the most suitable approaches for the chosen architecture and dataset are Self-Attention and Weighted methods for every available modality, and Self-Attention and Embracenet+ in the absence of a modality [15]. This paper introduces an innovative depiction of emotional state-capturing features extracted from user-generated Twitter data. The input representation is generated through the utilization of a sophisticated method based on the Genetic Algorithm (GA). This representation comprises linguistic, sentiment, and stylistic attributes that are extracted from tweets. Utilizing the novel feature representation, a voting ensemble classifier with weights optimized by a GA is presented in order to improve the precision of emotion detection [16]. The purpose of this research is to enhance social media sentiment analysis through the use of lengthier words. The novel lexicon-based system proposed by the researchers takes into account the lengthened word in its original form, rather than eliminating or normalizing it. Using framed lexicon constraints, the aggregated intensified senti-scores of lengthened words are computed and used to determine the individual's sentiment level. The study utilized a dataset consisting of informal Facebook conversations among various friend groups, Tweets, and personal messaging. In comparison to conventional systems that disregard lengthened terms, the proposed system achieves an 81% to 96% F-measure rate across all datasets, demonstrating its superior performance [17]. A transformer-based fusion and emotion-level representation learning approach is suggested in the study as a means to recognize emotions in multi-label videos. As inputs, the method receives text subtitles, raw video frames, and audio signals. Information from these multiple modalities is then passed through a unified transformer architecture in order to learn a joint multimodal representation [18]. The research paper provides a methodical examination of suggested methodologies, ranging from conventional to sophisticated, encompassing obstacles associated with sentiment analysis, feature engineering strategies, benchmark datasets, widely used publication platforms, and optimal algorithms for the progression of automated sentiment analysis [19]. The purpose of this study was to compare American, British, and Danish participants' perceptions of the emotional valence of affectively neutral text messages written in English [20]. By modifying the Multi-label K-Nearest Neighbors (MLkNN) classifier to take into account not only individual in-sentence features but also the features in neighboring sentences and the whole text of the tweet, this study attempts to enhance the accuracy and speed of emotion categorization for brief posts on Twitter. The enhanced L-MLkNN algorithm achieves a recall rate of 0.8019, which is higher than any of the competing approaches [21]. For the purpose of classifying emotions, they create deep learning models that are based on convolutional neural networks (CNN), long short-term memory (LSTM), bi-directional LSTM (BiLSTM), and the combination of convolutional neural networks and CLSTM. Class imbalance in the training dataset is addressed by using SMOTE and Random Undersampling strategies. With a 96% accuracy rate, the CNN model performed the best [22]. This study offers an ensemble method for tweet analysis in a variety of languages that makes use of a Generative Adversarial Network (GAN) and a Self-Attention Network (SAN). Enhancing the accuracy of sentiment analysis on a variety of textual data sets is the aim of the research project [23]. The paper discusses different approaches for feature extraction, including geometric-based, holistic, hybrid, and color-based methods. The paper mentions the use of data preprocessing techniques, such as cropping and scaling, to enhance the capabilities of deep learning [24].

III. PROPOSED FRAMEWORK

Combining modelling methodologies aims to produce a single model that analyses data from both text and face modalities simultaneously, enabling communication and information sharing between them. The goal of fusion approaches is to combine the features obtained from each method into a single representation that includes data from both text and facial images. For multi-modal emotion detection, integrating many modalities—such as text and facial images—requires the use of various techniques to efficiently combine and utilise data from each source. These methods seek to gather additional signals and provide a more thorough understanding of emotion. Figure 1 is showing proposed framework.

A. Preprocessing

Image preprocessing: We resize and grey-scale the photos in the face image dataset. We are extremely aware of the fact that deep networks perform well when dealing with huge datasets since learning from huge amounts of data teaches the deep neural network a lot about how to predict the feature very accurately, enabling it to perform well in subsequent predictions or classifications. In order to enhance the amount of the training data and the test data for the deep model, an augmentation process is done for data enhancement. With the goal of improving network input, visuals are normalized. To avoid overfitting and improve the network's resilience, the data set is further increased by randomly rotating, translating, and scaling the images.

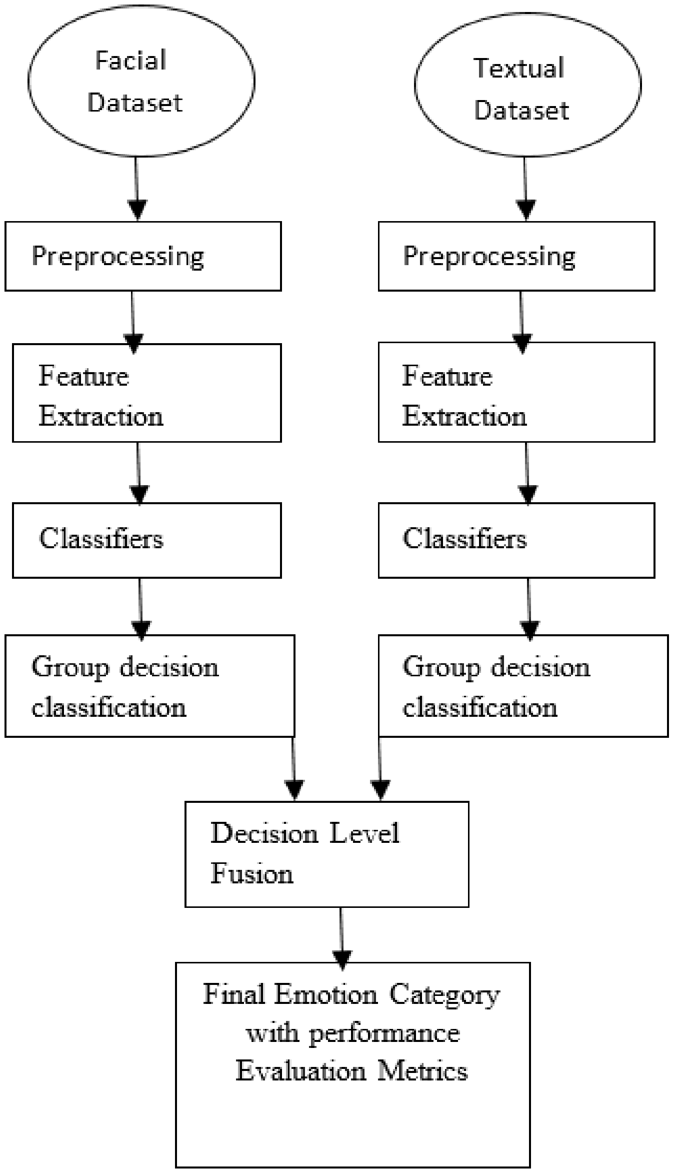


Figure 1: The Proposed Framework

Text Preprocessing: Text preparation is required to provide the data, which can substantially accelerate the sentiment analysis work. There are a variety of actions that can be taken to enhance text analysis tasks, such as switching to lowercase letters to avoid any duplication, eliminating stop-words, hyperlinks, and punctuation because they don't add any context for sentiment analysis, or spelling corrections because misspelled words can alter the sentiment of sentences, stemming to break down a word into its component parts by using a list of suffixes or prefixes, and lemmatization to maintain the meaning after keyword extraction unlike stemming. The smallest unit for natural language processing, such as a word or phrase, is created by tokenization in order to grasp the context and produce the best results. The steps for preprocessing text are shown in figure no. 2. Here addPartOfSpeech function is also applied after tokenization to retokenize the text for updating token details.

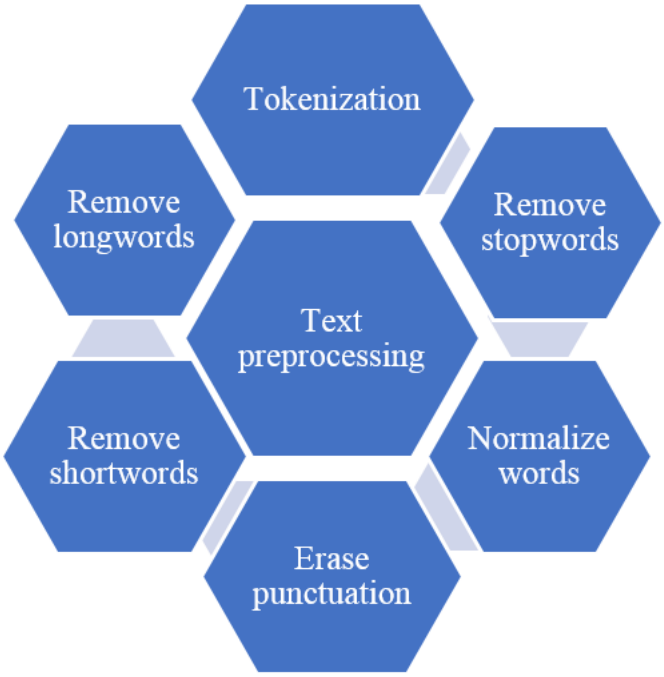


Figure 2: Text Preprocessing Diagram

B. Feature Extraction

VGG (Visual Geometry Group): VGG's deep architecture captures complex patterns in images. It performs well when a large amount of data is available for training. Its simple structure makes it easy to implement and modify and because of the pre-trained model, it saves training time. This model works very well for deep feature extraction of images, resulting in impressive accuracy rates in emotion recognition. This model excels at capturing both global and local patterns, enabling accurate classification even in complex scenarios.The convolutional part of VGG-16 consists of five sets of convolutional layers, each followed by max-pooling layers for spatial downsampling and the next 3 fully connected layers are utilized for feature extraction. The dataset contained images that were preprocessed as the VGG16 model works with the RGB color images with input size (224,224,3). As shown in the figure, to improve accuracy, features from face pictures are extracted into numerical vectors.

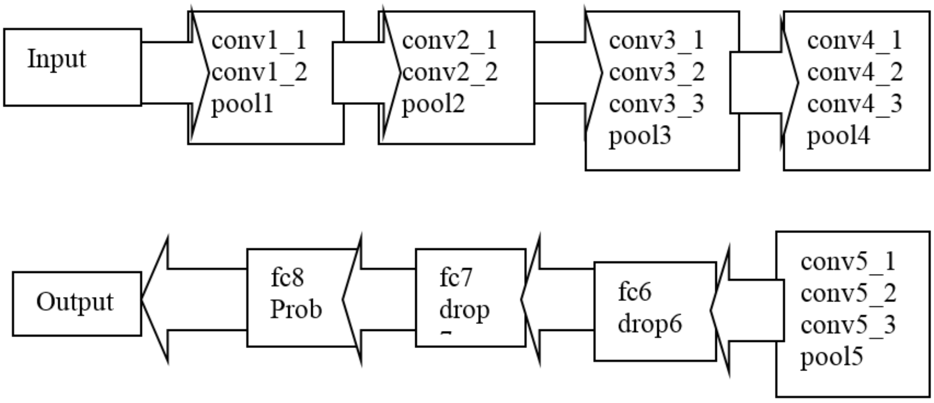


Figure 3: VGG 16 Architecture

Bag of Words (BoW): The bag of words technique makes it possible to handle and analyse vast volumes of text data efficiently by considering words as features. The text is preprocessed using the Bag of Words model, which turns it into a bag of words and counts how many times the most often used terms appear. It involves breaking down a piece of text into individual words and disregarding grammar and word order. The word-cloud figure shows the cleaned data after a bag of words preprocessing from raw data.

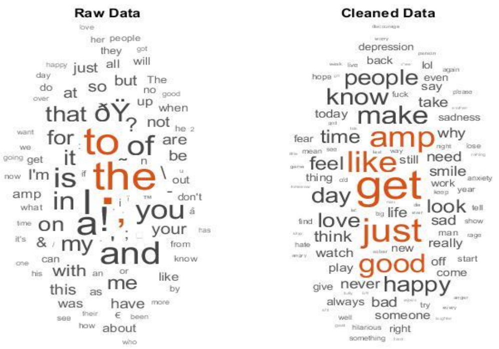


Figure 4: Raw and Cleaned Wordcloud Data Sample using BoW

C. Classification

SVMs are efficient at handling high-dimensional data and can capture complex decision boundaries. They are versatile and work well for both linear and nonlinear classification tasks. Decision tree (DT) are robust against overfitting and handle noisy data well. They can capture complex relationships in the data and provide feature importance, helping to understand model decisions. KNN is easy to understand and implement. It can capture local patterns and adapt to the structure of the dataset without assuming specific functional forms. An ensemble classifier is a machine-learning model that combines the predictions of multiple individual classifiers to make a final prediction. It leverages the diversity and complementary strengths of the individual classifiers to improve overall prediction accuracy and generalization performance.

Table I: Strengths and Limitations of Used Classifiers

Classifier	Strengths	Limitations	Applicability
K-Nearest Neighbors (KNN)	- Intuitive and easy to implement - Captures local patterns well - Suitable for multi-modal data integration	- Computationally expensive with large datasets - Sensitive to distance metric and K value selection - Struggles with high-dimensional data	Initial exploration, local pattern recognition
Support Vector Machines (SVM)	- Handles both linear and non-linear relationships - Effective in high-dimensional spaces - Fine-grained control over bias-variance trade-off	- Sensitive to hyperparameter and kernel choice - Training time and memory-intensive for large datasets - Requires well-structured data	High-dimensional, structured data, clear class boundaries
Decision Tree (DT)	- The automated feature interaction is supported by decision trees. Decision trees handle co-linearity better and are superior for categorical data.	It is unstable, hard to control tree size, susceptible to sampling mistakes, and offers a local ideal.	Branching occurs mostly by binary partitioning and the goal attributes are discrete or categorical.

D. Group Decision Classification

This research presents a group decision classifier system for textual and facial data sentiment analysis. It is built on four primary classifiers: decision trees (DT), KNNs, SVMs, and ensemble methods, in which the outputs of each classifier are merged using a majority voting mechanism. Majority voting is the most fundamental ensembled strategy for solving classification problems; the class with the most votes is selected. A voting classifier is a type of machine learning model that simply aggregates the outputs of all the classifiers it is fed, and uses the voting majority to predict the output class. Instead of creating separate, specialized models and evaluating each one's performance, the aim is to train a single model that forecasts output based on the cumulative majority of votes from each output class.

SVM-----Class1

KNN-----Class2

DT--------Class2 Group Decision Classification: Class label = mode {Class (joy), Class (sad), Class (fear), Class (anger)}

Ensemble--Class3

E. Decision Level Fusion

There are several methods for fusing textual and facial modalities together, including concatenating the features or using more advanced fusion strategies like early fusion, which combines the characteristics at a lower level, or late fusion, which combines the predictions of separate modality models. Late fusion, often referred to as decision-level fusion, entails processing each modality's properties individually before integrating the results at the point of final decision-making. This can entail averaging forecasts from various models or applying more sophisticated strategies like weighted averaging or stacking. To resolve mutual disambiguation across heterogeneous modalities, decision labels and probabilities from loosely connected modalities are modified using a technique called decision-level fusion. This technique depends on local judgments made by individual classifiers and avoids synchronization problems. When analysing problems in multi-criteria decision-making, aggregation functions like the arithmetic mean or average are utilized.

F. Final Emotion Prediction

Creating models that classify emotions can be useful for a number of purposes, such as sentiment analysis, customer feedback analysis, and computer-human interaction. Here, we effectively integrate the text and face modalities to extract emotion, ensuring that the detection of fine-grained sentiment analysis—that is, emotion—is unambiguous. Four sorts of emotions—joy, wrath, fear, and surprise—are the focus of this study.

IV. RESULTS AND DISCUSSIONS

A. Dataset

Visual Dataset: There are approximately 35,887 images in the FER2013 dataset. The dataset consists of grayscale images of different individuals, each displaying one of seven basic emotions: anger, disgust, fear, happiness, sadness, surprise, and neutral. FER2013 images are relatively small, typically 48x48 pixels.

The dataset originated from online sources, meaning that the images came from a variety of sources, leading to variations in terms of lighting, pose, and image quality. This diversity is both a strength and a challenge, as it helps the model generalize to different scenarios, but it also introduces noise and variability.

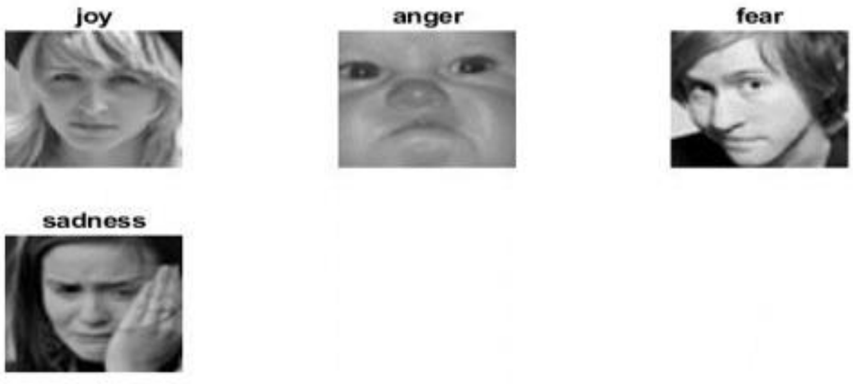


Figure 5: Facial Dataset Sample

Textual Dataset: Four different emotions—joy, sorrow, anger, and fear—are used to categorize tweets. A tweets dataset, Emotion Classification NLP, is used for textual data from the Kaggle website. In today's world use of social networking sites increasing consistently and a huge amount of data has been generated due to online community communication. Hence, we can use this data to determine the sentiment behavior of the user for many purposes like evaluation regarding any product, or service from comments, reviews, etc. A single sentence is analyzed as a standalone unit, and its overall tone is looked at. To analyze the emotions conveyed in tweets, a lot of studies are done using Twitter data. Being brief and casual, having mistakes in spelling, utilizing hashtags, unique symbols like emoticons and emojis, shortening words, and using acronyms are characteristics that set tweets apart from other types of communication and increase the difficulty of the task. Below example showing how the sentence-based document changed into tokens after tokenization.

{'You must be knowing #blithe means (adj.) Happy, cheerful.'}

{'Old saying 'A #smile shared is one gained for another day' @YEGlifer @Scott_McKeen'

After tokenized-document

6 tokens: know blithe mean adj happy cheerful

9 tokens: old say smile share gain another day yeglifer scottmckeen

B. Evaluation and Analysis

The percentage of similar occurrences that are involved in the retrieved occurrences is known as precision. Recall: The percentage of complimentary occurrences that have been extracted from the total number of occurrences is known as recall, often known as sensitivity. The comprehension and structure of relevance provide a sufficient foundation for both recall and accuracy. Recall and precision are integrated into the F-score, which is essentially the harmonic mean of the two metrics. The Decision Tree, K-NN, and Support Vector Machine models along with the ensemble learning model are evaluated in this study using four different metrics: precision, recall, F-measure, and accuracy.

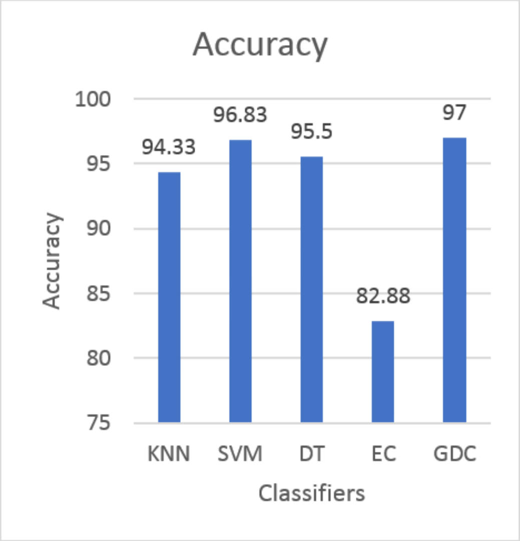
Accuracy=(TP+TN) / (TP+TN+FP+FN)

where TP stands for the total number of true categories that are both real and projected to be true; The amount of expected and actual incorrect categories is represented by the symbol TN;

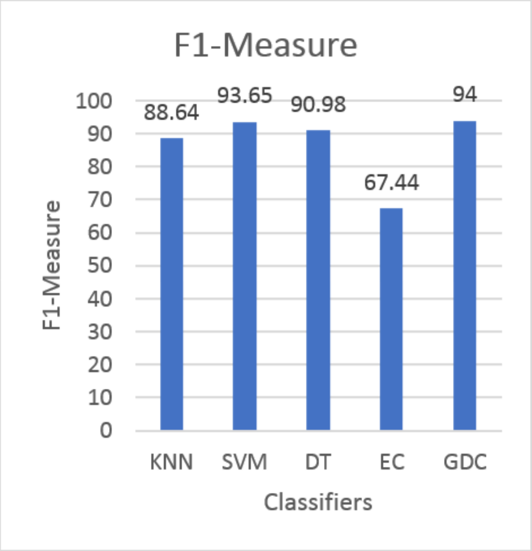
The number of actual incorrect categories and anticipated as true categories is shown by FN; FN indicates how many genuine true categories and those that were incorrectly predicted.

F1-Measure is calculated using the harmonic mean of precision and recall as follows:

F1-Measure=2*(Precision*Recall) / (Precision+Recall)



Graph 1: Accuracy Performance Evaluation



Graph 2: F1-Measure Performance Evaluation

The findings of the experiment from graph 1 and 2 confirm the effectiveness of our suggested framework technique, which is based on the group decision approach. It has the greatest accuracy when compared to four others widely used categorization methods that are already in use.

[25] Using the pre-trained networks MobileNetV2 and DenseNet-201, a novel feature-based transfer learning method is developed. The recognition rate of the proposed system is 75.31%. [26]The model demonstrates the exciting potential of ensemble-based approaches in facial emotion identification with an accuracy rate of 72.3%. [27]

This research uses computer vision, semantic recognition, and audio feature classification to analyse and identify the current types of human emotions in order to make artificial intelligence wiser by recognising user emotions. [28]

Here described model consists of three sub-networks, and the final result is obtained by integrating the output of the three networks using the SVM classifier. Using the FER2013 dataset, the model's expression's recognition accuracy was 71.27%.

Table II: Performance Evaluation Comparison With The Previous Research Studies

Authors	Significance of respective research method	Accuracy
Our proposed method (Overall accuracy on used of group decision classifier)	Group decision classification and decision level fusion of emotions using face and text	94%
[25]	DenseNet-201 and MobileNet-V2 are used for multi-network feature fusion, and KNN, NB, SVM, Ensemble with NCA, and augmentation are used for classification.	75.31%
[26]	ensemble-based categorization using ResNET50, InceptionV3, and Custom CNN	74.35%
[27]	Classification of facial expressions using an enhanced MobileNet model and feature fusion	72.3%
[28]	AlexNet, ResNet, and VGGNet, and SVM classifier	71.27%

V. CONCLUSION AND FUTURE WORK

In this study, we use the merging of text and face to accomplish autonomous emotion recognition. We evaluate a number of classifiers and decide to use the decision tree, SVMs, and k-nearest neighbor and ensemble classifiers. The categorization of text and face emotions has never been done previously using these classifier combinations. Along with decision-level fusion and group decision approach, we also put into practice data collecting, pre-processing, feature extraction, and the suggested classifier. This research has applications in boosting customer satisfaction, e-learning, psychological health care, and building smartphones and other gadgets that can sense human emotion. Finally, several links between the general sentiment of tweets and the major world events of the time can be discovered. Low generalization ability of the proposed system due to limited categories of discrete emotions. Since multiple emotions exist complexly in a single sentence, research on multi-labeled emotion recognition is needed using this specified approach. A multitask approach can constitute a viable solution for multimodal emotion recognition and sentiment analysis. The idea can be extended to combine feature vectorization methods and use other datasets to test the performance of this approach.

References

1) P. Nandwani and R. Verma, “A review on sentiment analysis and emotion detection from text,” Soc. Netw. Anal. Min., vol. 11, no. 1, pp. 1–19, 2021, doi: 10.1007/s13278-021-00776-6.

2) Y. Wang, J. Guo, C. Yuan, and B. Li, “Sentiment Analysis of Twitter Data,” Appl. Sci., vol. 12, no. 22, pp. 1–14, 2022, doi: 10.3390/app122211775.

3) A. Seyeditabari, N. Tabari, and W. Zadrozny, “Emotion Detection in Text: a Review,” 2018, [Online]. Available: http://arxiv.org/abs/1806.00674.

4) N. Perveen, N. Ahmad, M. A. Qadoos, B. Khan, R. Khalid, and S. Qadri, “Facial Expression Recognition Through Machine Learning,” Int. J. Sci. Technol. Res., vol. 5, no. 3, pp. 91–97, 2016.

5) F. A. Acheampong, C. Wenyu, and H. Nunoo-Mensah, “Text-based emotion detection: Advances, challenges, and opportunities,” Eng. Reports, vol. 2, no. 7, pp. 1–24, 2020, doi: 10.1002/eng2.12189.

6) A. Khlyzova, C. Silberer, and R. Klinger, “On the Complementarity of Images and Text for the Expression of Emotions in Social Media,” WASSA 2022 - 12th Work. Comput. Approaches to Subj. Sentim. Soc. Media Anal. Proc. Work., pp. 1–15, 2022, doi: 10.18653/v1/2022.wassa-1.1.

7) G. Yu, “Emotion Monitoring for Preschool Children Based on Face Recognition and Emotion Recognition Algorithms,” Complexity, vol. 2021, 2021, doi: 10.1155/2021/6654455.

8) B. Houshmand and N. M. Khan, “Facial Expression Recognition under Partial Occlusion from Virtual Reality Headsets based on Transfer Learning,” Proc. - 2020 IEEE 6th Int. Conf. Multimed. Big Data, BigMM 2020, pp. 70–75, 2020, doi: 10.1109/BigMM50055.2020.00020.

9) R. K. Singh, M. K. Sachan, and R. B. Patel, 360 Degree View of Cross-Domain Opinion Classification: a Survey, vol. 54, no. 2. Springer Netherlands, 2021.

10) O. Sagi and L. Rokach, “Ensemble learning: A survey,” Wiley Interdiscip. Rev. Data Min. Knowl. Discov., vol. 8, no. 4, pp. 1–18, 2018, doi: 10.1002/widm.1249.

11) R. Kaur and S. Kautish, “Multimodal sentiment analysis: A survey and comparison,” Int. J. Serv. Sci. Manag. Eng. Technol., vol. 10, no. 2, pp. 38–58, 2019, doi: 10.4018/IJSSMET.2019040103.

12) N. Ahmed, Z. Al Aghbari, and S. Girija, “A systematic survey on multimodal emotion recognition using learning algorithms,” Intell. Syst. with Appl., vol. 17, no. January, p. 200171, 2023, doi: 10.1016/j.iswa.2022.200171.

13) B. Kabra and C. Nagar, “Convolutional Neural Network based sentiment analysis with TF-IDF based vectorization,” J. Integr. Sci. Technol., vol. 11, no. 3, pp. 1–7, 2023.

14) C. L. Yang, S. Yoshida, H. Kuzuoka, T. Narumi, and N. Yamashita, “Affective Profile Pictures: Exploring the Effects of Changing Facial Expressions in Profile Pictures on Text-Based Communication,” Conf. Hum. Factors Comput. Syst. - Proc., 2023, doi: 10.1145/3544548.3581061.

15) D. Pena, A. Aguilera, I. Dongo, J. Heredia, and Y. Cardinale, “A Framework to Evaluate Fusion Methods for Multimodal Emotion Recognition,” IEEE Access, vol. 11, no. December 2022, pp. 10218–10237, 2023, doi: 10.1109/ACCESS.2023.3240420.

16) F. Anzum and M. L. Gavrilova, “Emotion Detection From Micro-Blogs Using Novel Input Representation,” IEEE Access, vol. 11, no. February, pp. 19512–19522, 2023, doi: 10.1109/ACCESS.2023.3248506.

17) A. Kukkar, R. Mohana, A. Sharma, A. Nayyar, and M. A. Shah, “Improving Sentiment Analysis in Social Media by Handling Lengthened Words,” IEEE Access, vol. 11, no. December 2022, pp. 9775–9788, 2023, doi: 10.1109/ACCESS.2023.3238366.

18) H. D. Le, G. S. Lee, S. H. Kim, S. Kim, and H. J. Yang, “Multi-Label Multimodal Emotion Recognition With Transformer-Based Fusion and Emotion-Level Representation Learning,” IEEE Access, vol. 11, no. December 2022, pp. 14742–14751, 2023, doi: 10.1109/ACCESS.2023.3244390.

19) D. Tiwari, B. Nagpal, B. S. Bhati, A. Mishra, and M. Kumar, A systematic review of social network sentiment analysis with comparative study of ensemble-based techniques, no. 0123456789. Springer Netherlands, 2023.

20) L. A. G. Neel, J. G. McKechnie, C. M. Robus, and C. J. Hand, “Emoji Alter the Perception of Emotion in Affectively Neutral Text messages,” J. Nonverbal Behav., vol. 47, no. 1, pp. 83–97, 2023, doi: 10.1007/s10919-022-00421-6.

21) X. Liu et al., “Emotion classification for short texts: an improved multi-label method,” Humanit. Soc. Sci. Commun., vol. 10, no. 1, pp. 1–9, 2023, doi: 10.1057/s41599-023-01816-6.

22) R. Olusegun, T. Oladunni, H. Audu, Y. A. O. Houkpati, and S. Bengesi, “Text Mining and Emotion Classification on Monkeypox Twitter Dataset: A Deep Learning-Natural Language Processing (NLP) Approach,” IEEE Access, vol. 11, no. May, pp. 49882–49894, 2023, doi: 10.1109/ACCESS.2023.3277868.

23) C. Kumaresan and P. Thangaraju, “ELSA: Ensemble learning based sentiment analysis for diversified text,” Meas. Sensors, vol. 25, no. December 2022, p. 100663, 2023, doi: 10.1016/j.measen.2022.100663.

24) T. Kumar Arora et al., “Optimal Facial Feature Based Emotional Recognition Using Deep Learning Algorithm,” Comput. Intell. Neurosci., vol. 2022, 2022, doi: 10.1155/2022/8379202.

25) V. Devi and A. Sharma, “Multi-Network Feature Fusion Facial Emotion Recognition using Nonparametric Method with Augmentation,” no. July, 2023.

26) Y. Chen and J. He, “Deep Learning-Based Emotion Detection,” J. Comput. Commun., vol. 10, no. 02, pp. 57–71, 2022, doi: 10.4236/jcc.2022.102005.

27) E. G. Moung, C. C. Wooi, M. M. Sufian, C. K. On, and J. A. Dargham, “Ensemble-based face expression recognition approach for image sentiment analysis,” Int. J. Electr. Comput. Eng., vol. 12, no. 3, pp. 2588–2600, 2022, doi: 10.11591/ijece.v12i3.pp2588-2600.

28) C. Jia, C. L. Li, and Z. Ying, “Facial expression recognition based on the ensemble learning of CNNs,” ICSPCC 2020 - IEEE Int. Conf. Signal Process. Commun. Comput. Proc., pp. 0–4, 2020, doi: 10.1109/ICSPCC50002.2020.9259543.

29) P. Nandwani and R. Verma, “A review on sentiment analysis and emotion detection from text,” Soc. Netw. Anal. Min., vol. 11, no. 1, pp. 1–19, 2021, doi: 10.1007/s13278-021-00776-6.

30) Y. Wang, J. Guo, C. Yuan, and B. Li, “Sentiment Analysis of Twitter Data,” Appl. Sci., vol. 12, no. 22, pp. 1–14, 2022, doi: 10.3390/app122211775.

31) A. Seyeditabari, N. Tabari, and W. Zadrozny, “Emotion Detection in Text: a Review,” 2018, [Online]. Available: http://arxiv.org/abs/1806.00674.

32) N. Perveen, N. Ahmad, M. A. Qadoos, B. Khan, R. Khalid, and S. Qadri, “Facial Expression Recognition Through Machine Learning,” Int. J. Sci. Technol. Res., vol. 5, no. 3, pp. 91–97, 2016.

33) F. A. Acheampong, C. Wenyu, and H. Nunoo-Mensah, “Text-based emotion detection: Advances, challenges, and opportunities,” Eng. Reports, vol. 2, no. 7, pp. 1–24, 2020, doi: 10.1002/eng2.12189.

34) A. Khlyzova, C. Silberer, and R. Klinger, “On the Complementarity of Images and Text for the Expression of Emotions in Social Media,” WASSA 2022 - 12th Work. Comput. Approaches to Subj. Sentim. Soc. Media Anal. Proc. Work., pp. 1–15, 2022, doi: 10.18653/v1/2022.wassa-1.1.

35) G. Yu, “Emotion Monitoring for Preschool Children Based on Face Recognition and Emotion Recognition Algorithms,” Complexity, vol. 2021, 2021, doi: 10.1155/2021/6654455.

36) B. Houshmand and N. M. Khan, “Facial Expression Recognition under Partial Occlusion from Virtual Reality Headsets based on Transfer Learning,” Proc. - 2020 IEEE 6th Int. Conf. Multimed. Big Data, BigMM 2020, pp. 70–75, 2020, doi: 10.1109/BigMM50055.2020.00020.

37) R. K. Singh, M. K. Sachan, and R. B. Patel, 360 Degree View of Cross-Domain Opinion Classification: a Survey, vol. 54, no. 2. Springer Netherlands, 2021.

38) O. Sagi and L. Rokach, “Ensemble learning: A survey,” Wiley Interdiscip. Rev. Data Min. Knowl. Discov., vol. 8, no. 4, pp. 1–18, 2018, doi: 10.1002/widm.1249.

39) R. Kaur and S. Kautish, “Multimodal sentiment analysis: A survey and comparison,” Int. J. Serv. Sci. Manag. Eng. Technol., vol. 10, no. 2, pp. 38–58, 2019, doi: 10.4018/IJSSMET.2019040103.

40) N. Ahmed, Z. Al Aghbari, and S. Girija, “A systematic survey on multimodal emotion recognition using learning algorithms,” Intell. Syst. with Appl., vol. 17, no. January, p. 200171, 2023, doi: 10.1016/j.iswa.2022.200171.

41) B. Kabra and C. Nagar, “Convolutional Neural Network based sentiment analysis with TF-IDF based vectorization,” J. Integr. Sci. Technol., vol. 11, no. 3, pp. 1–7, 2023.

42) C. L. Yang, S. Yoshida, H. Kuzuoka, T. Narumi, and N. Yamashita, “Affective Profile Pictures: Exploring the Effects of Changing Facial Expressions in Profile Pictures on Text-Based Communication,” Conf. Hum. Factors Comput. Syst. - Proc., 2023, doi: 10.1145/3544548.3581061.

43) D. Pena, A. Aguilera, I. Dongo, J. Heredia, and Y. Cardinale, “A Framework to Evaluate Fusion Methods for Multimodal Emotion Recognition,” IEEE Access, vol. 11, no. December 2022, pp. 10218–10237, 2023, doi: 10.1109/ACCESS.2023.3240420.

44) F. Anzum and M. L. Gavrilova, “Emotion Detection From Micro-Blogs Using Novel Input Representation,” IEEE Access, vol. 11, no. February, pp. 19512–19522, 2023, doi: 10.1109/ACCESS.2023.3248506.

45) A. Kukkar, R. Mohana, A. Sharma, A. Nayyar, and M. A. Shah, “Improving Sentiment Analysis in Social Media by Handling Lengthened Words,” IEEE Access, vol. 11, no. December 2022, pp. 9775–9788, 2023, doi: 10.1109/ACCESS.2023.3238366.

46) H. D. Le, G. S. Lee, S. H. Kim, S. Kim, and H. J. Yang, “Multi-Label Multimodal Emotion Recognition With Transformer-Based Fusion and Emotion-Level Representation Learning,” IEEE Access, vol. 11, no. December 2022, pp. 14742–14751, 2023, doi: 10.1109/ACCESS.2023.3244390.

47) D. Tiwari, B. Nagpal, B. S. Bhati, A. Mishra, and M. Kumar, A systematic review of social network sentiment analysis with comparative study of ensemble-based techniques, no. 0123456789. Springer Netherlands, 2023.

48) L. A. G. Neel, J. G. McKechnie, C. M. Robus, and C. J. Hand, “Emoji Alter the Perception of Emotion in Affectively Neutral Text messages,” J. Nonverbal Behav., vol. 47, no. 1, pp. 83–97, 2023, doi: 10.1007/s10919-022-00421-6.

49) X. Liu et al., “Emotion classification for short texts: an improved multi-label method,” Humanit. Soc. Sci. Commun., vol. 10, no. 1, pp. 1–9, 2023, doi: 10.1057/s41599-023-01816-6.

50) R. Olusegun, T. Oladunni, H. Audu, Y. A. O. Houkpati, and S. Bengesi, “Text Mining and Emotion Classification on Monkeypox Twitter Dataset: A Deep Learning-Natural Language Processing (NLP) Approach,” IEEE Access, vol. 11, no. May, pp. 49882–49894, 2023, doi: 10.1109/ACCESS.2023.3277868.

51) C. Kumaresan and P. Thangaraju, “ELSA: Ensemble learning based sentiment analysis for diversified text,” Meas. Sensors, vol. 25, no. December 2022, p. 100663, 2023, doi: 10.1016/j.measen.2022.100663.

52) T. Kumar Arora et al., “Optimal Facial Feature Based Emotional Recognition Using Deep Learning Algorithm,” Comput. Intell. Neurosci., vol. 2022, 2022, doi: 10.1155/2022/8379202.

53) V. Devi and A. Sharma, “Multi-Network Feature Fusion Facial Emotion Recognition using Nonparametric Method with Augmentation,” no. July, 2023.

54) Y. Chen and J. He, “Deep Learning-Based Emotion Detection,” J. Comput. Commun., vol. 10, no. 02, pp. 57–71, 2022, doi: 10.4236/jcc.2022.102005.

55) E. G. Moung, C. C. Wooi, M. M. Sufian, C. K. On, and J. A. Dargham, “Ensemble-based face expression recognition approach for image sentiment analysis,” Int. J. Electr. Comput. Eng., vol. 12, no. 3, pp. 2588–2600, 2022, doi: 10.11591/ijece.v12i3.pp2588-2600.

56) C. Jia, C. L. Li, and Z. Ying, “Facial expression recognition based on the ensemble learning of CNNs,” ICSPCC 2020 - IEEE Int. Conf. Signal Process. Commun. Comput. Proc., pp. 0–4, 2020, doi: 10.1109/ICSPCC50002.2020.9259543.