

# A Review of Geometric Deep Learning Technique for Emotional Detection Systems in the Context of Assistive Learning



M. O. Oloyede<sup>1</sup>, L. A. Olawoyin<sup>2</sup>, T. K. Hameed-Ibidokun<sup>3</sup>

<sup>1</sup>Department of Information Technology, University of Ilorin, Nigeria.

<sup>2</sup>Department of Telecommunication Science, University of Ilorin, Nigeria.

<sup>3</sup>NHS Business Service Authority, Newcastle, United Kingdom.

**ABSTRACT:** Emotional Detection Systems (EDS) are critical for understanding human reactions in human-computer interaction, yet accurately identifying emotions from facial expressions, text, and speech remains challenging. While Geometric Deep Learning (GDL) has emerged as a promising paradigm for processing non-Euclidean data structures, its application in affective computing remains underexplored. This paper presents a critical review of EDS modalities, techniques, and datasets, with particular emphasis on assessing the integration of GDL approaches. A literature analysis across IEEE Xplore, Science Direct, and Springer reveals significant research disparity, with facial expression being the most extensively studied modality. The review identifies key challenges including individual variability in emotional expression, difficulty inferring complex internal states such as culture and intent, under-represented emotion categories, and insufficient dataset diversity. An examination of existing facial, text, and speech datasets further highlights the absence of comprehensive emotion type coverage. Findings indicate that GDL adoption in emotion detection remains partial, with limited studies leveraging its capabilities and significant room for maturation in graph-construction practices. This paper concludes that future progress requires integrating GDL's structural intelligence with multimodal, interpretable, and privacy-preserving paradigms to develop adaptive, human-centric systems capable of perceiving and responding to emotional states with improved accuracy, empathy, and ethical responsibility.

**KEYWORDS:** Emotional detection, geometric deep learning, emotion datasets, facial expression, speech emotion, text emotion.

[Received Mar. 10, 2025; Revised Aug. 10, 2025; Accepted Dec. 08, 2025]

Print ISSN: 0189-9546 | Online ISSN: 2437-2110

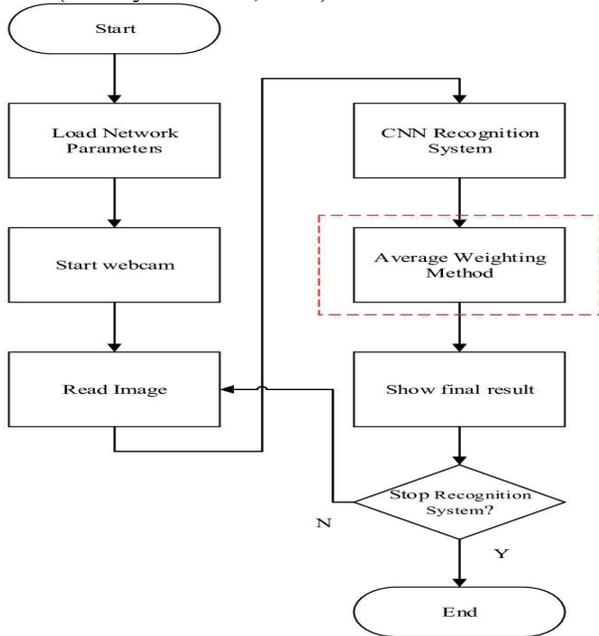
## I. INTRODUCTION

Emotions play a crucial role in daily life, influencing decision-making, communication, and situational awareness in human-centric environments. As a result, they have become a significant area of interest across various research fields (Krishnamoorthy et al., 2024; Yang et al., 2021). Emotional Detection Systems (EDS) have gained increasing attention due to their importance in human-computer interaction, helping to analyze and understand human behavioral traits. The rise of emotional computing has further driven research in this area, enabling the development of software that can interpret human emotional states and respond accordingly (Singh & Gill, 2022). Nandwani et al. (Nandwani & Verma, 2021) define EDS as systems capable of identifying distinct human emotions, such as happy, depression, or anger. Despite the research efforts on Natural Language Processing (NLP), the approaches face critical limitations because text-based systems often misinterpret subtle emotional cues or mixed sentiments, particularly in informal or social media language. The absence of structural knowledge and multimodal integration is another disadvantage. The relational or graph-like structures that underpin emotional meaning, such as relationships between entities, phrases, and speech patterns, are ignored by text-based emotion models, which view language as a straight line of words. This limits the models' capacity to represent complex emotions through deeper semantic relationships. Additionally,

the majority of review papers focus on algorithmic comparisons while ignoring practical and ethical issues like dataset bias, imbalance among emotion categories, and the cultural subjectivity of emotional labeling. Reproducibility and real-world adaptation are further restricted by data paucity, inconsistent annotation, and overfitting to benchmark datasets. Lastly, there is a research void for further investigation because few works examine how Geometric Deep Learning can describe the non-linear, relational elements of text emotion structures.

The ability to recognize emotions is a key focus in machine learning research, particularly in assistive computing, where it supports applications in emotion engineering (Lim et al., 2020; Oloyede et al., 2018). Additionally, emotion detection plays a vital role in interpersonal communication, with emotional expressions varying across different cultural contexts (Liu et al., 2024). Although data subjectivity, multimodality issues, cultural prejudice, and ethical concerns limit its efficacy, emotion detection for disinformation improves understanding of how emotions contribute to the development and dissemination of erroneous information. In order to capture the relationship structure of emotional and informational signals inside social networks, future research should focus on interpretable, cross-domain, and multimodal emotion models, maybe using Geometric Deep Learning (GDL). EDS can be implemented using various modalities,

such as shown in Figure 1. These modalities make emotional detection applicable in a wide range of fields, such as neuroscience, human-computer interaction, psychology, music recommendation systems, e-learning, and cognitive science (Murthy & Kumar, 2021).

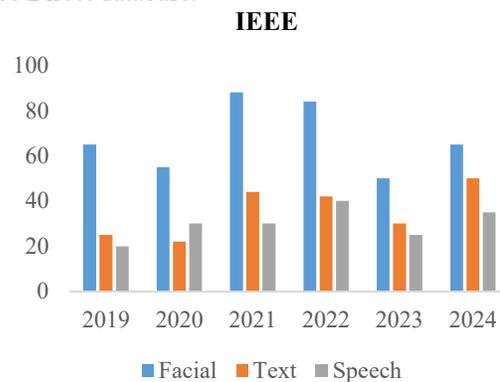


**Figure 1** Modalities of Emotion Detection System

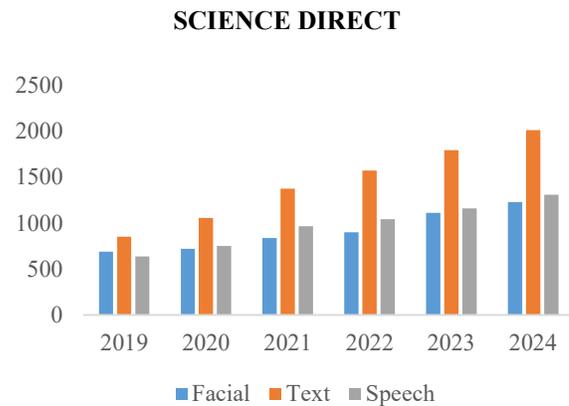
For example, text-based emotion detection allows teachers to understand students' emotions, enabling them to adapt educational resources and activities to better support learning within a given educational community (Bustos-López et al., 2021). According to Sreeja, P & Mahalakshmi, G., (2017), text reflects the emotional state of both the writer and the reader and can be analysed using various computational models. Similarly, speech-based emotion detection extracts feature from speech to identify underlying emotions. Zamil et al., (2019) highlighted that speech features capture information related to emotion, accent, and rhythm, which are influenced by factors such as vocal cord tension, pressure, and distance (Jain et al., 2018; Tariq et al., 2019). Facial expression-based emotion detection has numerous applications, including surveillance systems, gaming, and entertainment (Khan et al., 2019). Chen et. al., (2021) showed that evaluating viewers' emotions through facial expression analysis during online video viewing can improve the viewing experience. In their investigation, the authors concluded that the novelty is in intent-based adaptation, in which the system determines the purpose of a user's viewing and adjusts network resource allocation accordingly. The suggested methodology bridges the gap between Quality of Experience (QoE), measurement and real user intent by integrating emotion detection, user context modeling, and AI-driven resource management to match network decisions with human-centric satisfaction measures. Nevertheless, consistent datasets and criteria that integrate network-level, contextual, and emotional data for thorough QoE assessment are lacking. These drawbacks

emphasize the need for more reliable, scalable, and privacy-preserving frameworks that can strike a compromise between technological efficiency and user-centred experience.

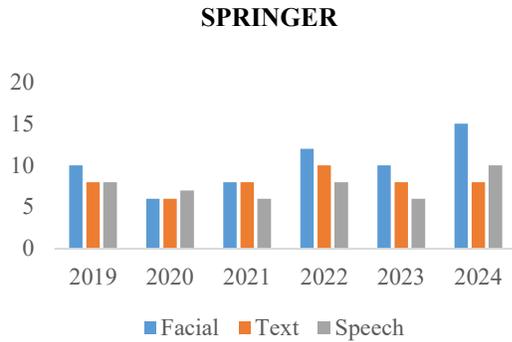
The extent of research focus on different emotion detection modalities remains unclear in the literature. To investigate this, a search was conducted across three academic databases i.e. IEEE Xplore, Science Direct, and Springer using the keywords “Emotion detection by facial expression,” “Emotion detection by text,” and “Emotion detection by speech.” The goal was to analyse research trends and determine which modalities have received the most attention over the past six years. As shown in Figures 2–4, the search results indicate that emotion detection through facial expressions has been the most widely studied modality in the IEEE Xplore and Springer databases. In contrast, emotion detection through text has been the primary focus in the Science Direct database.



**Figure 2** Graph showing the research disparity in the different modalities of emotion on the IEEE database.



**Figure 3** Graph showing the research disparity in the different modalities of emotion on the Science Direct database.



**Figure 4** Graph showing research disparity in the different modalities of emotion on the Springer database.

Various techniques and approaches have been proposed for developing emotion detection systems across different modalities. However, according to the literature, many of these techniques fail to achieve optimal performance (Alsubai, 2023). Additionally, despite their effectiveness in machine learning applications, Geometric Deep Learning (GDL) techniques have yet to be widely applied in emotion detection systems (Gerken et al., 2021). GDL refers to a set of emerging techniques that extend deep neural models to non-Euclidean domains, such as graphs and manifolds (Shagdar et al., 2021). This study explores the different techniques used in developing emotion detection systems, with a particular focus on GDL techniques, as discussed in Section II. Unlike previous research, which typically focuses on a single modality such as facial expressions, text, or speech, this study provides a comprehensive survey of all three. Additionally, it examines the potential of GDL techniques in the context of emotion detection. Hence, the key contributions of this study are as follows:

- i. A survey of three primary modalities for emotion detection i.e. facial expressions, text, and speech.
- ii. An analysis of various techniques used in developing emotion detection systems.
- iii. A detailed examination of GDL techniques and their potential applications in emotion detection.
- iv. A presentation of existing emotion detection datasets, along with a review of performance results reported by researchers.

The remainder of this paper is organized as follows: Section II explores the various modalities used in emotion detection systems and examines how researchers have utilized these modalities in their developments. Section III provides a brief overview of Geometric Deep Learning (GDL) and its key concepts. Section IV discusses the application of GDL in emotion detection systems. Section V reviews different emotion detection datasets, highlighting their limitations. Section VI summarizes the key insights from this study and outlines directions for future research. Section VII presents the conclusion of the study.

## II MODALITIES OF EMOTION DETECTION SYSTEMS

Human decision-making and communication are greatly influenced by emotions. Emotion detection (ED) methods have thus grown in importance as a field of study, especially when it comes to human-machine interactions. Emotion detection systems (EDS) are being developed in both industry and academia due to the increasing importance of ED (Garcia-Garcia et al., 2017). These systems have a wide range of applications, including education, neuroscience, cognitive science, artificial intelligence, and computer science (Binali & Potdar, 2012). Given their diverse applications, researchers are continually working to develop more effective ED techniques. Since EDS is inherently human-centric, its techniques are derived from various human characteristics, referred to as sources in this study. Research indicates that 7% of an individual's communicated information is conveyed through posture, 38% through speech, 30% through facial expressions, and 25% through text (Oloyede et al., 2020). These different sources provide valuable insights into human emotions. Although many studies in the literature primarily focus on facial expressions when developing ED techniques, this section explores alternative sources, specifically text and speech. The following subsections examine how researchers have utilized these modalities to develop emotion detection techniques.

Emotion detection systems use a variety of modalities, including speech, text, facial expressions, and physiological signs, to automatically identify human emotions. The intricate, non-Euclidean correlations present in real-world emotional cues, such as spatial dependencies between face landmarks or brain connectivity in EEG signals, are frequently missed by conventional deep learning models like CNNs and RNNs, despite their effectiveness for structured data. By extending deep learning to the graph and manifold domains, Geometric Deep Learning (GDL) overcomes this constraint by allowing models to handle irregular data structures and comprehend relational information between features. In this area, methods such as Graph Convolutional Networks (GCNs), Graph Attention Networks (GATs), and Message Passing Neural Networks (MPNNs) are frequently employed (Farhad et. Al., 2021).

By applying GDL across multiple modalities, researchers achieve richer and more context-aware emotion detection. Facial graphs can represent landmark connections, speech can be modeled as temporal graphs of acoustic features, and EEG data can form brain connectivity networks. When combined, multimodal geometric fusion captures cross-modal emotional dependencies for higher accuracy and robustness. Despite challenges such as graph construction complexity, limited labeled data, and computational cost, GDL-based emotion systems show great promise for applications in affective computing, mental health monitoring, and human-computer interaction. Future directions include dynamic graph learning, self-supervised GDL, and federated geometric models for privacy-preserving, real-time emotion recognition (Garcia-Garcia, et.al., 2017).

### A. Facial Expression Modality

The facial expression modality is one of the most widely used approaches in EDS because the human face naturally reflects emotional states through muscle movements and expressions. This involves capturing facial data by using images, videos, or 3D facial meshes and analysing features such as eyebrow position, eye openness, mouth curvature, and cheek movement to infer emotions like happy, anger, sad, fear, disgust, surprise, and neutrality. EDS is facial expression modality typically uses Convolutional Neural Networks (CNNs) for spatial feature extraction, 3D-CNNs or LSTMs for temporal (video) emotion tracking, and increasingly Graph Convolutional Networks (GCNs) to model geometric relationships between facial landmarks. This allows the system to understand subtle muscle changes and spatial dependencies across different facial regions. Facial EDSs are non-invasive and effective, but they face challenges such as lighting variations, head pose changes, occlusion, and cultural differences in expression. Despite these, it remains a core modality for emotion detection due to its direct visual interpretability and compatibility with real-time systems.

Different facial expressions of individuals can show whether their faces are static or in motion (Oloyede et al., 2020). These expressions serve as key indicators of emotions and mental states, playing a crucial role in daily emotional communication (Tarnowski et al., 2017). While humans can naturally recognize facial expressions, this ability has been leveraged in automatic emotion detection systems to improve emotion recognition (Verma & Verma, 2020). Since an individual's face is an exposed part of the body, computer vision systems can analyse facial images from photos or videos for emotion detection and recognition (Zhong et al., 2017). To enhance emotion detection, researchers have developed various emotion recognition systems. Tarnowski et al., (2017) utilized a Microsoft Kinect for 3D face modelling due to less cost and ease of operation. The Kinect has limited scanning resolution and a high rate of registering (30 frames), with an infrared emitter with dual cameras. One of the cameras records visible light, and the other functions as infrared used for measuring depth. A 3D face model is created when the infrared rays are reflected on the user's body. The model is centred on

121 specific points of the face, recorded by the Kinect device. Spatial coordinates of these points are stored in the form of a matrix, and changes in facial expression derived from the movement of facial muscles have been taken into consideration. Experiments for seven emotional states carried out on the KDEF database showed 86% accuracy for random division of data; and 73% for natural division of data.

In (Oloyede et al., 2019), the author earlier proposed a face recognition system that incorporated metaheuristic algorithms for image enhancement during pre-processing. This approach increased the number of detectable features in facial images, improving classification performance when integrated with a convolutional neural network (CNN). Experiments conducted on the Yale Face dataset, specifically addressing the effect of facial expressions, achieved 93.8% accuracy using their method. Samara et al., (2019), focused on affective state recognition based on facial expressions during computerized tasks. Data was collected from users interacting with software while completing predefined tasks. A machine learning approach was then applied to analyze facial expressions. However, the authors found that facial expressions alone were insufficient to accurately reflect users' emotions in common computerized interactions, as experiments on the CK-8 and KDEF datasets yielded unsatisfactory results.

Also, in (Turabzadeh et al., 2018), researchers developed a real-time facial expression recognition system, tested in MATLAB and later in MATLAB Simulink. Their system identified facial expressions at one frame per second. The dataset used for evaluation consisted of videos recorded from five individuals watching a video, and the approach produced satisfactory results. Minaee et al., (2021), proposed a deep learning-based emotion detection technique using an attentional convolutional neural network (CNN). Their model focused on the most relevant facial regions, improving performance on datasets such as FERG, JAFFE, and FER-2013. However, the study also revealed that different emotions are associated with different facial regions, making some emotions more challenging to detect than others. Table 1 show the modality of facial expression while Table 2 further summarizes the different approaches that have been utilized in developing emotion detection systems via facial expression.

**Table 1** The table showing Facial Expression Modality

| Modality          | Data Representation                  | Feature Extraction                          | Common Models               | Emotion Output             |
|-------------------|--------------------------------------|---|-----------------------------|----------------------------|
| Facial Expression | Image landmarks, 3D meshes           | Pixel intensity, geometry temporal dynamics | CNN, GCN, 3D-CNN, LSTM      | Basic/compound emotion     |
| Text              | Word embeddings, syntactic graphs    | Semantic, lexical. Contextual features      | LSTM, BERT, GNN             | Emotion or sentiment class |
| Speech            | Acoustic signals, spectrogram, MFCCs | Pitch, energy, tone, prosody                | CNN, LSTM, GAT, Transformer | Affective state            |

**Table 2** Summary of different emotion detection systems developed via facial expression modality

| Authors                 | Year | Dataset   | Emotions                  | Comment   |
|-------------------------|------|---|---------------------------|---|
| Tarnowski et al., 2017  | 2017 | KDEF  | Sad, and Happy            | Utilized a Microsoft Kinect for 3D face modeling due to less cost and ease of operation, where a 3D face model is created when the infrared rays are reflected on the user's body.  |
| Oloyede et al., 2019    | 2019 | Yale  | Sad, Happy, and Surprise. | Effective image enhancement method using metaheuristic algorithms for pre-processing face images with expression. Introduction of hybrid features to improve the classification performance of convolutional neural network architecture. |
| Samara et al., 2019     | 2019 | CK-8  | Happy, and Smile.         | The machine learning method is used for facial expression-based affective state recognition. Experiments did not show satisfactory results.   |
| Turabzadeh et al., 2018 | 2018 | Data collected from five individuals watching a video | Smile, Anger              | Facial expression real-time system that could identify facial expressions expression in real-time at a rate of 1 frame per second.  |
| Minaee et al., 2021     | 2021 | JAFFE   | Smile, Surprise           | The deep learning technology is based on an attentional convolutional neural network that could focus on vital sections of the face.  |

## B. Text Modality

In recent years, written text has evolved into various forms, including social media posts, news articles, and microblogs (Sailunaz et al., 2018). These textual contents serve as valuable resources for text mining, enabling the analysis of emotions in contexts such as insulting language, suicide notes, and multimedia tagging (Al Maruf et al., 2024). Text-based emotion detection (ED) focuses on identifying the emotions that drive individuals to write certain sentences at specific times (Acheampong et al., 2020). However, extracting emotions from text is a complex task, prompting researchers in both industry and academia to develop more effective approaches for text-based emotion detection (Tu et al., 2024). As a result, numerous methods have been proposed, leveraging word- and sentence-based techniques, natural language processing (NLP), and machine learning algorithms (Zad et al., 2021). Azmin & Dhar, (2019), authors proposed a Multinomial Naïve Bayes (BN) based emotion detection model from Bangla text which is one of the top seven most spoken languages in the world. The detection method was divided into four sequential phases i.e. dataset preparation, preprocessing, feature selection and extraction, and emotion classification respectively. The dataset for the analysis was collected from different Facebook groups and posts of popular bloggers.

The dataset was preprocessed and categorized into three emotion classes: happy, sad, and angry. Feature selection and extraction involved a combination of part-of-speech (POS) tagging, word n-grams, and term frequency-inverse document frequency (TF-IDF) vectorization. Experimental results showed that using bigram-based TF-IDF combined with POS features, stop words, and an emotion-based processor yielded the highest accuracy, achieving 78.6% overall. In (Tuhin et al., 2019), the authors proposed an emotion detection model using two text-mining-based algorithms: Naïve Bayes and a topical approach. Their dataset, consisting of 7,500 Bangla sentences,

was categorized into six emotion classes: happy, sad, tender, excited, angry, and scared. A comparative analysis revealed that the topical approach outperformed Naïve Bayes, achieving the best overall performance. Hasan et al., (2019) proposed an automatic emotion detection method based on a supervised learning approach. The approach was divided into two main tasks: an offline training activity and an online classification task. The offline task involves Emotex, which collects large emotion-labeled messages as a dataset from Twitter, pre-processed, and extracts feature vectors to train the emotion classification model. The online task involves utilizing the created model to classify the live streams of tweets for emotion detection in real-time. The classification of emotion data was done with Naïve Bayes (NB), Support Vector Machine (SVM) and Decision Tree (DT).

To solve the Fuzzy boundaries nature of automatic emotion detection, the authors employed the dimensional emotion model and soft fuzzy classification approach where the latter produced better results. In (Zhong et al., 2019), the authors proposed a Knowledge Enriched Transformer (KET), which utilizes self-attention to interpret contextual utterances while dynamically incorporating external knowledge through a context-aware affective graph attention mechanism. Experimental results on multiple textual datasets demonstrated that integrating both text and external knowledge significantly enhances the performance of emotion detection systems. Zhang et al., (2018), introduced an emotion distribution learning approach and developed a multi-task convolutional neural network (CNN) for text-based emotion detection. Their model effectively learns robust representations from distributional datasets annotated by multiple voters. The end-to-end framework enhances both prediction and classification tasks simultaneously, leading to improved performance. Table 3 further summarizes the different approaches that have been utilized in developing emotion detection systems via text.

**Table 3** Summary of different emotion detection systems developed via text modality

| Authors            | Year | Dataset                     | Emotions  | Comment  |
|--------------------|------|-----------------------------|---|--|
| Azmin & Dhar, 2019 | 2019 | Facebook groups and Blogs.  | Happy, Sad, and Angry.                          | Naïve Bayes (BN) based emotion detection model from Bangla text. The experiment showed that combining bigram-based tf-idf with POS features, stop word, and emotion-based processor produced better performance.   |
| Tuhin et al., 2019 | 2019 | Balanga Text Dataset        | Happy, Sad, Tender, Excited, Angry, and Scared. | The emotion detection model is based on two text-mining-based algorithms namely, Naïve Bayes and topical approach. The experiment showed that the topical approach achieved better performance.  |
| Hasan et al., 2019 | 2018 | Data collected from Twitter | Sad and Happy                                   | Effective emotion detection method based on supervised learning approach. The approach was divided into two main tasks: an offline training activity and an online classification task. The tasks could collect data from tweets and classify online tweets for emotion detection. |
| Zhong et al., 2019 | 2019 | Textual Dataset             | Happy, Sad, and Surprise                        | Utterances are interpreted utilizing self-attention and external ideas. The authors concluded that both text and ideas are needed to improve emotion detection systems.  |
| Zhang et al., 2018 | 2018 | ISEAR                       | Happy and Sad                                   | A multi-task convolutional neural network that can learn robust representation for the distribution dataset with annotations of various voters.  |

### C. Speech Modality

Speech is one of the most natural and effective modalities for human expression and has gained significant attention in emotion detection research (Akçay & Oğuz, 2020). It enables the rapid exchange of ideas and emotions, making it a valuable source for identifying an individual's emotional state (Jain et al., 2018). Consequently, integrating speech-based emotion detection into computer applications has become an important area of study (Luna-Jiménez et al., 2021). Jain et al., (2018) proposed a speech emotion detection system that involves feature extraction and utilizes a Cubic Spline Support Vector Machine (SVM) for classification. Their experiments, conducted on Hindi-language speech data, demonstrated satisfactory performance in emotion detection.

In (Zamil et al., 2019), the authors developed a speech-based emotion detection system by extracting Mel Frequency Cepstrum Coefficient (MFCC) features from speech signals. They then applied a Logistic Model Tree classification approach to categorize the emotions. Experiments on the Berlin Database of Emotional Speech (Emo-DB) and the Ryerson Audio-Visual Database of Emotional Speech showed promising results for detecting certain emotions. Additionally, in (Dhiman et al., 2021), the authors introduced a speech emotion detection system using modified dense convolutional neural networks (CNNs). Their method leveraged paralinguistic speech features, such as vocal tract characteristics, and classified emotions using spectrograms of audio files. Experiments conducted on the Emo-DB and SAVEE emotional datasets demonstrated strong performance in recognizing emotional states.

Furthermore, Tariq et al., (2019) proposed a real-time speech detection system based on the Internet of Things (IoT) and deep learning, specifically designed for elderly individuals in care facilities. To enhance classification performance, the system incorporated data normalization and augmentation techniques. Experiments conducted using the Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) demonstrated satisfactory performance in detecting emotions such as calm, sad, angry, fearful, and disgust. Tripathi et al., (2019), developed a speech-based emotion detection system that integrates both speech features and textual transcriptions. Spectrograms and Mel-Frequency Cepstral Coefficients (MFCC) were utilized to extract low-level speech characteristics related to emotions, while transcriptions helped capture semantic meanings.

The combination of these features improved emotion detection. Experiments conducted on the University of Southern California's Interactive Emotional Motion Capture (USC-IEMOCAP) dataset showed a slight performance improvement over state-of-the-art techniques in recognizing emotions such as anger, neutral, sad, and happy. In Krishna et al., (2022), focused on detecting emotions conveyed by speakers during speech. Their approach employed various classification algorithms, including Support Vector Machines (SVM), Multi-Layer Perceptron (MLP), and Tonnetz features. These models were trained to recognize emotions such as happy, sad, surprise, and fear. Their method achieved an accuracy of 85% in experimental evaluations. Table 4 further summarizes the different approaches that have been utilized in developing emotion detection systems via speech.

**Table 4** Summary of different emotion detection systems developed via speech.

| Authors               | Year | Dataset  | Emotions                               | Comment   |
|-----------------------|------|--|--|---|
| Jain et al., 2018     | 2018 | Hindi language speech Dataset  | Happy and Sad                          | The system constitutes the extraction of unique features and utilizes a cubic spine support vector machine and the classification method.   |
| Zamil et al., 2019    | 2019 | Berlin Database of Emotional Speech                                      | Surprise, Anger, and Happy             | Extraction of Mel Frequency Cepstrum Coefficient (MFCC) features from speech signals, and the Logistic Model Tree classification approach was used to classify the features.  |
| Dhiman et al., 2021   | 2021 | Berlin Database of Emotional Speech                                      | Sad and Happy                          | Paralinguistic features of speech like vocal tract feature and performs classification of speech-based emotions using spectrograms of its audio files.  |
| Tariq et al., 2019    | 2019 | Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS)     | Calm, Sad, Angry, Fearful, and Disgust | Internet of Things (IoT) based deep learning system for old adults in care facilities.  |
| Tripathi et al., 2019 | 2019 | University of Southern California's Interactive Emotional Motion Capture | Anger, Neutral, Sad, and Happy         | Spectrogram and Mel-frequency Cepstral Coefficients (MFCC) are used as speech features and assist in selecting emotion-related low-level characteristics in speech, while transcriptions assist in capturing semantic meanings. |
| Krishna et al., 2022  | 2023 | Hindi language speech Dataset  | Happy, Sad, Surprise, and Fearful      | Different classification algorithms such as support vector machine, multi-layer perceptron, and Tonnetz were utilized in detecting series of emotions.  |

### III GEOMETRIC DEEP LEARNING

Deep learning approaches have achieved significant success in machine learning applications such as speech recognition, image classification, and object detection (Cao et al., 2020). In particular, Convolutional Neural Networks (CNNs), unlike traditional neural networks, leverage the fundamental statistical properties of data, including local stationarity and multi-scale component structures, to extract features and capture deep local information. While deep learning has been highly effective in processing structured data such as video, text, and audio, most research has primarily focused on data defined in the Euclidean domain, i.e., grid-like structures (Monti et al., 2019). However, with the availability of larger datasets and advancements in GPU computing power, researchers have shifted their attention to processing data in non-Euclidean domains, such as manifolds and graphs, a field known as Geometric Deep Learning (GDL) (Koch et al., 2019).

GDL techniques are designed to handle graph-structured data, which consist of nodes and edges representing network structures. Additionally, they process manifold data, often used to describe geometric shapes such as object surfaces (Atz et al., 2021). The geometric data are not uniformly organized, thus making it a complex task for people to identify underlying patterns and define convolution processes on images. Whereas data like images in the Euclidean domain are known as special graph data with vertices organized in a uniform manner (Koch et al., 2019). In addition, another problem in non-Euclidean data is that it often has numerous large scales, making it

difficult to utilize a deep learning approach to process analysis and classification tasks. Hence, deep learning is needed in geometric data, where emotion detection problems fall in this category (Cao et al., 2020).

### IV APPLICATION OF GDL ON EMOTION DETECTION SYSTEMS

In this section, we highlight recent and relevant research where the Geometric Deep Learning (GDL) approach has been applied, particularly in the development of emotion detection systems. A data-driven deep learning approach was proposed in (Besson et al., 2021) to analyse brain shape without relying on manually defined features. Their method builds on the principles of geometric deep learning by adapting a standard Convolutional Neural Network (CNN) architecture to the surface representation of the cortical ribbon. For their experiment, data from 6,410 healthy subjects were extracted from 11 publicly available datasets. Their model achieved an 87.99% accuracy in predicting a subject's sex and a correlation coefficient of 0.93 with an absolute error of 4.58 years when estimating a subject's age. In (James, 2021), a geometric deep learning approach was introduced for citywide traffic speed prediction.

This method captures geometric dependencies in traffic data using graph convolution processes, while temporal data correlations are extracted and refined through an encoder-decoder architecture within a generative adversarial learning framework. Experiments conducted on real-world urban road networks demonstrated improved performance compared to

other approaches. To predict the structure organization of the human visual cortex, Ribeiro et al., (2021) designed a geometric deep learning-based system. The system can exploit the structure of the cortex to understand the relationship between brain function and structure in the human visual cortex such that important maps can be predicted. Their approach can also predict different emotions amongst individuals, thereby showing the approach could be applied to different applications involving data structured in non-Euclidean spaces.

To address the challenge of accurately estimating a well-centred and representative Connectional Brain Template (CBT) for populations of multi-view brain networks (MVBNs), Gurbuz & Rekik, (2020) proposed a geometric deep learning architecture called *Deep Graph Normalizer*. This model normalizes a population of MVBNs by integrating them into a unified CBT. Their approach learns to merge MVBNs while capturing non-linear variations within subjects, all while preserving the topological properties of brain graphs through the use of Graph CNNs. Experiments conducted on both small- and large-scale datasets demonstrated superior performance compared to existing methods.

To explore the application of Geometric Deep Learning (GDL) techniques in the development of emotion detection systems, a literature search was conducted across three academic databases. Findings revealed that limited technical research has been dedicated to applying GDL for emotion detection. However, a few studies have utilized GDL in this context, as outlined below. Authors in (Nguyen et al., 2021) the authors proposed a GDL-based approach called *PointNet++*, applied to 3D point cloud data for emotion detection. Their method involved a preprocessing step that included face center cropping and point cloud denoising on 3D face scans. To optimize performance, a hyper parameter tuning process was conducted.

Experiments on the Bosphorus dataset achieved accuracy rates of 69.01% and 85.85% for recognizing seven and five different facial expressions, respectively. In (Ahmed et al., 2021), a GDL-based approach was introduced to develop a text-based hateful comment detection system that incorporates social network information. The inclusion of network context improved performance compared to other techniques including support vector machine, linear regression, and artificial neural network. Also, the performance metrics used in the study include accuracy, precision, recall, and F1-Score. Liu et al., (2021) proposed a GDL based facial expression recognition system that integrates a Spatial Attention Convolutional Neural Network (CNN) with a series of Long Short-Term Memory (LSTM) networks enhanced by an Attention Mechanism. Additionally, they introduced a deep geometric feature descriptor to capture the relative geometric position correlations of facial landmarks. Experimental results on multiple facial expression datasets demonstrated that their approach outperformed existing techniques.

## V. EMOTION DETECTION DATASET

This section presents three major emotion detection datasets based on different modalities: facial expressions, text, and speech. These datasets include one or more emotional categories and are widely used by researchers to validate various approaches and techniques.

### A. Facial expression-based Emotion Detection Datasets

The following sub-sections describe the different emotion datasets based on facial expressions. A summary of the facial datasets and the emotion types present is shown in Table 5. Furthermore, some performance results on each dataset are presented.

**Table 5** Emotion types present in Facial Expression Based Emotion Detection Dataset

| Face Dataset  | Different Emotion Types |       |         |       |     |          |         |         |
|---------------|-------------------------|-------|---------|-------|-----|----------|---------|---------|
|               | Smile                   | Anger | Fearful | Happy | Sad | Surprise | Disgust | Neutral |
| JACFEE        | ✓                       | ✓     |         | ✓     | ,   | ✓        | ✓       | ✓       |
| CMU Multi-PIE | ✓                       | ✓     | ✓       |       |     | ✓        |         | ✓       |
| AffectNet     | ✓                       |       | ✓       | ✓     | ,   | ✓        | ✓       | ✓       |

#### 1) Japanese and Caucasian Facial Expressions of Emotion

The Japanese and Caucasian Facial Expressions of Emotion (JACFEE) is a facial dataset that consists of 56 coloured static pictures of 56 different persons who depict one of the seven various emotions. The emotions in this dataset include anger, contempt, disgust, fear, happy, sad, and surprise. The gathering of the pictures is evenly divided into

28 males and females and contains equal numbers of Caucasian and Japanese individuals i.e. 14 Caucasian males, 14 Caucasian females, 14 Japanese males, and 14 Japanese females. In addition, the JACFEE dataset comes with a document that describes the history of the JACFEE images and the percent judgment of intended emotion for each image. Table 6 shows performance from different research on the JACFEE face dataset.

**Table 6** Results performance from different research on the JACFEE dataset

| Research work               | Technique/Approach   | Emotions Considered                       | Emotion Recognition Performance  |
|-----------------------------|--|---|--|
| Castro-Vale et al., 2015    | A log linear model was utilized to confirm various emotions on war veterans and their respective offspring. The model further confirmed if the agreement between the proposed and the chosen emotions was higher than expected.  | Sad and Happy.                            | Sad – 63%<br>Happy – 91%   |
| Benitez-Garcia et al., 2018 | For the analysis of Western-Caucasian facial expression, the visual analysis leveraged the Eigenfaces technique, and the cross-cultural facial expression recognition combines geometric and appearance features by getting facial Fourier descriptors and local Fourier coefficients. | Anger, Disgust, Happy, Sad, and Surprise. | Anger – 89.5%<br>Disgust – 91%<br>Fear – 88.5%<br>Happy – 94.5%<br>Sad – 87%<br>Surprise – 94.5% |
| Susskind et al., 2007       | Individual Support Vector Machines were trained to develop effective recognition of one of six basic emotional expressions in real-time with no explicit training on expression similarity.  | Happy, Surprise, Sad, Anger, and Disgust. | Happy – 89.2%<br>Surprise – 92.9%<br>Sad – 91.8%   |

### 2) CMU Multi-PIE

The CMU Multi-PIE dataset includes 75,000 face images of 337 individuals recorded in four different sessions over five months. Subjects are captured in different poses and under

illumination conditions while showing different emotions via facial expressions such as Happy, Relaxed, Sad, and Stressed. Table 7 shows performance from different research on the CMU Multi-PIE face dataset.

**Table 7** Results performance from different research on the CMU Multi-PIE dataset

| Research work                 | Technique/Approach   | Emotions Considered                             | Emotion Recognition Performance  |
|-------------------------------|--|---|--|
| Chen et al., 2021             | An innovative deep learning model utilizing the kernel entropy component analysis network at the feature extraction stage and the directed acyclic graph support vector machine as a classifier for the expression recognition.  | Smile, Surprise, Disgust, Scared, and Neutral.  | Smile – 80.02%<br>Surprise – 92.44%<br>Disgust – 85.90%<br>Scared – 93.23%<br>Neutral – 84.72%             |
| Du & Hu, 2019                 | Integration of Manifold Regularization Dictionary Pair Learning and Intra-class Variation Reduced Feature model with Iterative Optimization Classification Strategy. The approach systematically eradicates non-useful samples to get an effective representation model. | Anger, Disgust, Fear, Happy, Sad, and Surprise. | Anger – 88.54%<br>Disgust – 92.03%<br>Fear – 94.56%<br>Happy – 88.70%<br>Sad – 86.76%<br>Surprise – 96.82% |
| Kuruvayil & Palaniswamy, 2021 | Meta-learning utilizes prototypical networks to solve the issue of lack of enough samples for facial emotion recognition. The approach enables the use of a few training samples.  | Disgust, Happy, Neutral, Anger, and Surprise.   | Disgust – 80%<br>Happy – 77%<br>Neutral – 87%<br>Anger – 93%<br>Surprise – 100%                            |

### 3) AffectNet

The AffectNet facial dataset contains over one million images from the internet by querying three search engines with 1250 emotion-related words in different languages. Up to half of the facial images were annotated to consider different facial

expressions i.e. Happy, Sad, Neutral, Surprise, Fear, and Disgust. AffectNet is considered the largest dataset for facial expression recognition (Mollahosseini et al., 2017). Table 8 shows performance from different research on the AffectNet Dataset.

**Table 8** Results performance from different research on the AffectNet dataset

| Research work              | Technique/Approach   | Emotions Considered                                      | Emotion Recognition Performance  |
|----------------------------|--|--|--|
| Mollahosseini et al., 2017 | Two models based on deep neural networks are utilized to classify facial images in the categorical model and assume the intensity of different emotions.   | Neutral, Happy, Sad, Surprise, Fear, Disgust, and Anger. | Neutral – 50.8%<br>Happy – 70%,<br>Sad – 69.7%,<br>Surprise – 66.5%,<br>Fear – 61.1%,<br>Disgust – 67.6%,<br>Anger – 62.3%.    |
| Li et al., 2021            | Pure transformer-based mask vision transformer to get a mask that can filter out backgrounds and occluded faces. Also, the approach contains a dynamic relabeling module to correct labels in the face datasets in the wild.   | Surprise, Fear, Disgust, Happy, Sad, Anger, and Neutral. | Surprise – 63.80%<br>Fear – 61.20%<br>Disgust – 55.00%<br>Happy – 86.80%<br>Sad – 63.80%<br>Anger – 59.20%<br>Neutral – 62.20% |
| Zhao et al., 2021          | Global multi-scale and local network that contains a feature pre-extractor, multi-scale module, and a local attention module. All three components have individual functions such as to extract middle-level features, fusing features with different receptive fields, and guiding the network to concentrate on local salient features respectively. | Different facial expressions.                            | 64.53%   |

**B. Text-Based Emotion Detection Datasets**

The different emotion datasets based on text are presented. A summary of the text-based emotion datasets and

the emotion type present is shown in Table 9. Furthermore, some performance results on each dataset are presented.

**Table 9** Emotion types present in the Text-Based Emotion Detection Datasets

| Text-based Emotion Dataset  | Different Emotion Types |       |         |       |     |          |         |         |
|---|-------------------------|-------|---------|-------|-----|----------|---------|---------|
|   | Smile                   | Anger | Fearful | Happy | Sad | Surprise | Disgust | Neutral |
| Kaggle  | ✓                       | ✓     |         |       |     |          | ✓       | ✓       |
| The International Survey on Emotion Antecedents and Reactions (ISEAR) | ✓                       | ✓     | ✓       | ✓     | ✓   |          |         | ✓       |
| GoEmotions  |                         | ✓     |         |       |     | ✓        | ✓       | ✓       |

**1) Kaggle**

The Kaggle text emotion dataset collects tweets annotated with different emotions behind them. It consists of three columns i.e. tweet\_id, sentiment, and content that

contains the raw tweet. While the sentiment includes the emotion that comes with the tweet Table 10 shows the result performance from different research on the Kaggle Dataset.

**Table 10** Results performance from different research on the Kaggle Dataset

| Research work         | Technique/Approach  | Emotions Considered       | Emotion Recognition Performance                      |
|-----------------------|---|---------------------------|--|
| Mehta et al., 2020    | The deep learning-based model incorporates standard psycholinguistic features and language model embedding to assume emotions.        | Anger                     | Anger – 60.04%                                       |
| Kumar et al., 2019    | Binary adaptation of Cuckoo search is used to select features for emotion analysis from textual online contents.                      | Smile, Disgust, and Anger | Smile – 70.08%<br>Disgust – 65.80%<br>Anger – 74.50% |
| Rupapara et al., 2021 | An ensemble-based approach named regression vector voting classifier was utilized to detect angered comments on social media outlets. | Anger                     | Anger – 90%  |

2) **ISEAR**

The International Survey on Emotion Antecedents and Reactions (ISEAR) dataset was collected by asking students in different disciplines to report scenarios in which they had experienced seven types of emotions i.e. joy, fear, anger, sad,

disgust, shame, and guilt. The question covered how they appraised the situation and how they reacted. The ISEAR dataset contains reports on the seven emotions by 3000 respondents in 37 countries on all continents. Table 11 shows performance from different research on the ISEAR Dataset

**Table 11** Results performance from different research on the ISEAR Dataset

| Research work       | Technique/Approach  | Emotions Considered               | Emotion Recognition Performance                                    |
|---------------------|---|-----------------------------------|--|
| Kumar & Raman, 2022 | Dual-channel approach for multi-class text emotion recognition that includes an embedding module, dual-channel module, emotion, and classification module.                    | Happy, Sad, Hate, and Anger.      | Happy – 82%,<br>Sad – 79%,<br>Hate – 75%,<br>Anger – 80%           |
| Hasan et al., 2021  | An effective sequential transfer learning approach named DeepEmotex to detect emotion in text.  | Joy, Sad, and Anger.              | Joy – 60%,<br>Sad – 58%<br>Anger – 56%                             |
| Alotaibi, 2019      | Supervised learning approach utilizing Linear regression classifier to solve the issue of limited coverage of emotional clue that resulted in less performance of the system. | Joy, Fear, Sad, Shame, and Guilt. | Joy – 70%<br>Fear – 62%<br>Sad – 73%<br>Shame – 70%<br>Guilt – 58% |

3) **GoEmotions**

The GoEmotions text emotion dataset contains 58,000 English Reddit comments that have been manually annotated and labeled for different emotion categories such

as love, joy, anger, fear, confusion, disgust, desire, gratitude, remorse, pride, grief, embarrassment, and surprise. Table 12 shows performance from different research on the GoEmotions Dataset.

**Table 12** Results performance from different research on the ISEAR Dataset

| Research work                 | Technique/Approach  | Emotions Considered             | Emotion Recognition Performance                          |
|-------------------------------|---|---------------------------------|--|
| Olah et al., 2021             | Transfer of knowledge gained from the supervision on social media tweets utilizing various emotion representation approaches.   | Joy, Disgust, and Anger.        | Joy – 60.4%<br>Disgust – 56.5%<br>Anger - 65.3%          |
| Alvarez-Gonzalez et al., 2021 | Developed a benchmark and investigated several feature spaces with learning algorithms. The differences between how users show emotions and how they comprehend them is also studied. | Anger, Fear, Sad, and Surprise. | Anger – 50%<br>Fear – 56%<br>Sad – 48%<br>Surprise – 60% |
| Cortiz, 2021                  | Applying Natural Language Understanding by fine-tuning various transformers language models and evaluating them in terms of performance.  | Anger, Fear, Joy, and Love.     | Anger – 47%<br>Fear – 37%<br>Joy – 53%<br>Love – 78%     |

C. **Speech-Based Emotion Detection Datasets**

The following sub-sections describe the different emotion datasets based on Speech. A summary of the

speech-based emotion datasets and the emotion types present is shown in Table 13. Furthermore, some performance results on each dataset are presented.

**Table 13** Emotion types present in the Speech-Based Emotion Detection Dataset

| Speech-based Emotion Dataset                      | Different Emotion Types |       |         |       |     |          |         |         |
|---|-------------------------|-------|---------|-------|-----|----------|---------|---------|
|   | Smile                   | Angry | Fearful | Happy | Sad | Surprise | Disgust | Neutral |
| URDU-Dataset                                      |                         | ✓     |         | ✓     |     |          |         | ✓       |
| Keio University Japanese Emotional Speech Dataset |                         | ✓     |         |       | ✓   |          | ✓       | ✓       |
| Multimodal EmotionLines Dataset                   |                         | ✓     | ✓       |       | ✓   | ✓        | ✓       | ✓       |

**1) URDU-Dataset**

The URDU dataset includes emotional speeches collected from Urdu talk shows. It includes 400 utterances of three basic emotions i.e. Angry, Happy, and Neutral, and

consists of 38 speakers (27 male and 11 female). Table 14 shows performance from different research on the URDU-Dataset.

**Table 14** Results performance from different research on the URDU-Dataset.

| Research work       | Technique/Approach  | Emotions Considered | Emotion Recognition Performance |
|---------------------|---|---------------------|---------------------------------|
| Latif et al., 2018  | Features that are frame-level and knowledge-inspired are extracted using the openSMILE toolkit. This toolkit allows the extraction of numerous audio features. The Support Vector Machine is used as the feature classification method. | Anger and Happy     | Anger – 60.4%<br>Happy – 50.8%  |
| Naeem et al., 2020  | Improvements in the architecture of a statistical automatic speech recognition system. Also, the Kaldi toolkit is used for training the Urdu speech recognition system.   | Anger and Happy.    | Anger – 75.6%<br>Happy – 80.4%  |
| Farhad et al., 2021 | Pitch and Cepstral features are collected from audio files, and principal component analysis is used for the reduction of dimensionality. In addition, Random Forest is utilized as the classifier.                                     | Sad and Happy.      | Sad – 72.2%<br>Happy – 74.5%    |

**2) Keio University Japanese Emotional Speech Database**

The Keio University Japanese Database was created in 2006 and consists of human speech with male vocal emotion

spoken by Japanese. It includes emotions such as anger, joy, disgust, funny, and calm. Table 15 shows performance from different research on the Keio University Japanese Emotional Speech Dataset.

**Table 15** Results performance from different research on the Keio University Japanese Emotional Speech Dataset

| Research work         | Technique/Approach   | Emotions Considered      | Emotion Recognition Performance                 |
|-----------------------|--|--------------------------|---|
| Tomba et al., 2018    | The Mel-Frequency Cepstral Coefficient approach is used as the classification method for an emotion recognition system.            | Disgust and Sad.         | Disgust – 60%<br>Sad – 80%.                     |
| Moritani et al., 2021 | Application of StarGAN voice conversion to an emotional voice conversion system with reduced fundamental frequency and processing. | Sad, Anger, and Neutral. | Sad – 73.3%<br>Anger – 43.3%<br>Neutral – 63.3% |
| Suzuki et al., 2021   | Implemented a face to voice emotion recognition system using a 3D model.   | Sad, Happy, and Disgust. | Sad – 56.6%<br>Happy – 60.5%<br>Disgust – 68.7% |

**3) Multimodal EmotionLines Dataset**

The Multimodal EmotionLines Dataset (MELD) is created to improve and extend the EmotionLines dataset. It consists of above 1400 dialogues and 13000 words from the

media, where multiple speakers attended the dialogues. Every dialogue is labelled with emotions such as anger, disgust, sad, joy, fear, surprise, and neutral. Table 16 shows performance from different research on the Multimodal EmotionLines Dataset.

**Table 16** Results performance from different research on the Multimodal EmotionLines Dataset

| Research work   | Technique/Approach  | Emotions Considered                      | Emotion Recognition Performance  |
|-----------------|---|--|--|
| Ho et al., 2020 | Multimodal-based speech emotion recognition using Multi-Level Multi-Head Fusion Attention mechanism and recurrent neural network. | Neutral, Surprise, Fear, Sad, and Anger. | Neutral – 78.6%<br>Surprise – 56.9%<br>Fear – 38.46%<br>Sad – 28.12%<br>Anger – 57.04% |

|                       |   |   |   |
|-----------------------|---|---|---|
| Choi et al.,<br>2021  | A residual-based graph convolution network that contains an intra-utterance feature extractor that investigates the intra-inter informative features. | Angry, Sad, and Happy.                    | Angry – 69.75%<br>Sad – 76.17%<br>Happy – 91.63%                                  |
| Zhang et al.,<br>2019 | A conventional graph-based convolutional neural network that depicts every utterance and speaker as a node.   | Anger, Disgust, Fear, Joy, Fear, and Sad. | Anger – 46.8%<br>Disgust – 10.6%<br>Fear – 8.7%<br>Joy – 53.1%<br>Neutral – 76.7% |

## VI LESSONS LEARNT / FUTURE RESEARCH DIRECTION

In the preceding sections, we have presented extensive research on emotion detection systems; and the various modalities in which emotion detection systems can be developed. We also investigated the application of Geometric Deep Learning (GDL) techniques in emotion detection applications. It is observed that few research has utilized geometric deep learning to develop emotion detection systems. Hence it is recommended that the GDL approach should be utilized in the area of emotion detection systems due to its ability to analyse and classify complex tasks. Furthermore, the different emotion detection datasets based on the various modalities are presented, and it is observed that most of these datasets do not include some basic emotion types. Thus, making it difficult for researchers to come up with systems that can detect certain emotions. Following the critical review of this study, this section presents the following as future research directions. Modalities of Emotion Detection Systems: In this study, it is clear that the major source of emotion includes facial expression, text, and speech. Also, from the analysis in figures 2 – 4, it is evident that facial expression is the most researched and utilized in developing emotion detection systems. However, other modalities of emotion detection systems such as posture, can be investigated and researched to confirm their performance in detecting emotions.

Geometric Deep Learning: Geometric Deep Learning is an effective approach to different applications because it can process data in a non-Euclidean domain like manifolds and graphs. However, from this study, it is clear that it has not been explored in emotion detection and affective

computing. Hence, researchers in this field can explore the GDL approach in applying Emotion Detection Systems, as present techniques and approaches have not reached innovative performance. Emotion Detection Datasets: the emotion detection datasets are presented based on the different modalities and emotion types. It is observed that not all the emotion types are present in the different selected datasets. Hence, emotion detection datasets must be developed to consider more emotion types.

## VII CONCLUSION

In this study, we reviewed academic literature on emotion detection systems and geometric deep learning. Also, recently published articles were reviewed to appreciate the research work that has been done in the field of emotion detection systems. In this regard, we have presented in figures 2-4, the disparity in research on the different modalities of emotions in three different academic databases including IEEE, Science direct, and Springer. We also presented the different modalities of emotion detection i.e. facial expression, text, and speech. Different techniques and approaches based on these different modalities were presented and analysed. Furthermore, a summary of the different approaches based on the various emotion detection modalities has been presented. The emotion detection dataset that contains the different emotion types is discussed, and emotion detection performance is presented. It is observed that most of these datasets do not include some basic emotion types. Thus, making it difficult for researchers to come up with systems that can detect certain emotions.

## REFERENCES

**Acheampong, F. A., C. Wenyu, & H. Nunoo-Mensah.** (2020). Text-based emotion detection: Advances, challenges, and opportunities. *Engineering Reports*, 2(7), e12189.

**Ahmed, Z., B. Vidgen, & S. A. Hale.** (2021). Tackling racial bias in automated online hate detection: Towards fair and accurate classification of hateful online users using geometric deep learning. *arXiv preprint arXiv:2103.11806*.

**Akçay, M. B., & K. Oğuz.** (2020). Speech emotion recognition: Emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers. *Speech Communication*, 116, 56–76.

**Al Maruf, A., F. Khanam, M. M. Haque, Z. M. Jiyad, F. Mridha, & Z. Aung.** (2024). Challenges and opportunities of text-based emotion detection: A survey. *IEEE Access*.

- Alotaibi, F. M.** (2019). Classifying text-based emotions using logistic regression. *VAWKUM Transactions on Computer Sciences*, 7(1), 31-37.
- Alsubai, S.** (2023). Emotion Detection Using Deep Normalized Attention-Based Neural Network and Modified-Random Forest. *Sensors*, 23(1), 225.
- Alvarez-Gonzalez, N., A. Kaltenbrunner, & V. Gómez.** (2021). Uncovering the limits of text-based emotion detection. arXiv preprint arXiv:2109.01900.
- Atz, K., F. Grisoni, & G. Schneider.** (2021). Geometric deep learning on molecular representations. *Nature Machine Intelligence*, 1-10.
- Azmin, S., & K. Dhar.** (2019). Emotion detection from Bangla text corpus using Naïve Bayes classifier. 2019 4th International Conference on Electrical Information and Communication Technology (EICT).
- Benitez-Garcia, G., T. Nakamura, & M. Kaneko.** (2018). Multicultural facial expression recognition based on differences of western-caucasian and east-asian facial expressions of emotions. *IEICE TRANSACTIONS on Information and Systems*, 101(5), 1317-1324.
- Besson, P., T. Parrish, A. K. Katsaggelos, & S. K. Bandt.** (2021). Geometric deep learning on brain shape predicts sex and age. *Computerized Medical Imaging and Graphics*, 91, 101939.
- Binali, H., & V. Potdar.** (2012). Emotion detection state of the art. *Proceedings of the CUBE International Information Technology Conference*.
- Bustos-López, M., N. Cruz-Ramírez, A. Guerra-Hernández, L. N. Sánchez-Morales, & G. Alor-Hernández.** (2021). Emotion Detection from Text in Learning Environments: A Review. *New Perspectives on Enterprise Decision-Making Applying Artificial Intelligence Techniques*, 483-508.
- Cao, W., Z. Yan, Z. He, & Z. He.** (2020). A comprehensive survey on geometric deep learning. *IEEE Access*, 8, 35929-35949.
- Castro-Vale, I., M. Severo, D. Carvalho, & R. Mota-Cardoso.** (2015). Emotion recognition ability test using JACFEE photos: a validity/reliability study of a war veterans' sample and their offspring. *Plos one*, 10(7), e0132293.
- Chen, X., L. Ke, Q. Du, J. Li, & X. Ding.** (2021). Facial expression recognition using kernel entropy component analysis network and DAGSVM. *Complexity*, 2021.
- Choi, Y.-J., Y.-W. Lee, & B.-G. Kim.** (2021). Residual-based graph convolutional network for emotion recognition in conversation for smart Internet of Things. *Big Data*, 9(4), 279-288.
- Cortiz, D.** (2021). Exploring transformers in emotion recognition: a comparison of bert, distillbert, roberta, xlnet and electra. arXiv preprint arXiv:2104.02041.
- Dhiman, R., G. S. Kang, & V. Gupta.** (2021). Modified dense convolutional networks based emotion detection from speech using its paralinguistic features. *Multimedia Tools and Applications*, 80(21), 32041-32069.
- Du, L., & H. Hu.** (2019). Weighted patch-based manifold regularization dictionary pair learning model for facial expression recognition using iterative optimization classification strategy. *Computer Vision and Image Understanding*, 186, 13-24.
- Farhad, M., H. Ismail, S. Harous, M. M. Masud, & A. Beg.** (2021). Analysis of Emotion Recognition from Cross-lingual Speech: Arabic, English, and Urdu. 2021 2nd International Conference on Computation, Automation and Knowledge Management (ICCAKM).
- Garcia-Garcia, J. M., V. M. Penichet, & M. D. Lozano.** (2017). Emotion detection: a technology review. *Proceedings of the XVIII international conference on human computer interaction*.
- Gerken, J. E., J. Aronsson, O. Carlsson, H. Linander, F. Ohlsson, C. Petersson, & D. Persson.** (2021). Geometric deep learning and equivariant neural networks. arXiv preprint arXiv:2105.13926.
- Gurbuz, M. B., & I. Rekik.** (2020). Deep graph normalizer: a geometric deep learning approach for estimating connectional brain templates. *International Conference on Medical Image Computing and Computer-Assisted Intervention*.
- Hasan, M., E. Rundensteiner, & E. Agu.** (2019). Automatic emotion detection in text streams by analyzing twitter data. *International Journal of Data Science and Analytics*, 7(1), 35-51.
- Hasan, M., E. Rundensteiner, & E. Agu.** (2021). DeepEmotex: Classifying Emotion in Text Messages using Deep Transfer Learning. 2021 IEEE International Conference on Big Data (Big Data).
- Ho, N.-H., H.-J. Yang, S.-H. Kim, & G. Lee.** (2020). Multimodal approach of speech emotion recognition using multi-level multi-head fusion attention-based recurrent neural network. *IEEE Access*, 8, 61672-61686.
- Jain, U., K. Nathani, N. Ruban, A. N. J. Raj, Z. Zhuang, & V. G. Mahesh.** (2018). Cubic SVM classifier based feature extraction and emotion detection from speech signals. 2018 International Conference on Sensor Networks and Signal Processing (SNSP).
- James, J.** (2021). Citywide traffic speed prediction: A geometric deep learning approach. *Knowledge-Based Systems*, 212, 106592.
- Khan, G., A. Siddiqi, M. U. G. Khan, S. Q. Wahla, & S. Samyan.** (2019). Geometric positions and optical flow based emotion detection using MLP and reduced dimensions. *IET Image Processing*, 13(4), 634-643.
- Koch, S., A. Matveev, Z. Jiang, F. Williams, A. Artemov, E. Burnaev, M. Alexa, D. Zorin, & D. Panozzo.** (2019). Abc: A big cad model dataset for geometric deep learning. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
- Krishna, K. V., N. Sainath, & A. M. Posonia.** (2022). Speech Emotion Recognition using Machine Learning.

2022 6th International Conference on Computing Methodologies and Communication (ICCMC),

**Krishnamoorthy, P., M. Sathiyarayanan, & H. P. Proença.** (2024). A novel and secured email classification and emotion detection using hybrid deep neural network. *International Journal of Cognitive Computing in Engineering*, 5, 44–57.

**Kumar, A., A. Jaiswal, S. Garg, S. Verma, & S. Kumar.** (2019). Sentiment analysis using cuckoo search for optimized feature selection on Kaggle tweets. *International Journal of Information Retrieval Research (IJRR)*, 9(1), 1–15.

**Kumar, P., & B. Raman.** (2022). A BERT based dual-channel explainable text emotion recognition system. *Neural Networks*.

**Kuruvayil, S., & S. Palaniswamy.** (2021). Emotion recognition from facial images with simultaneous occlusion, pose and illumination variations using meta-learning. *Journal of King Saud University-Computer and Information Sciences*.

**Latif, S., A. Qayyum, M. Usman, & J. Qadir.** (2018). Cross lingual speech emotion recognition: Urdu vs. western languages. 2018 International Conference on Frontiers of Information Technology (FIT).

**Li, H., M. Sui, F. Zhao, Z. Zha, & F. Wu.** (2021). Mvt: Mask vision transformer for facial expression recognition in the wild. arXiv preprint arXiv:2106.04520.

**Lim, J. Z., J. Mountstephens, & J. Teo.** (2020). Emotion recognition using eye-tracking: taxonomy, review and current challenges. *Sensors*, 20(8), 2384.

**Liu, C., K. Hirota, J. Ma, Z. Jia, & Y. Dai.** (2021). Facial expression recognition using hybrid features of pixel and geometry. *IEEE Access*, 9, 18876–18889.

**Liu, Z., T. Zhang, K. Yang, P. Thompson, Z. Yu, & S. Ananiadou.** (2024). Emotion detection for misinformation: A review. *Information Fusion*, 102300.

**Luna-Jiménez, C., R. Kleinlein, D. Griol, Z. Callejas, J. M. Montero, & F. Fernández-Martínez.** (2021). A Proposal for Multimodal Emotion Recognition Using Aural Transformers and Action Units on RAVDESS Dataset. *Applied Sciences*, 12(1), 327.

**Mehta, Y., S. Fatehi, A. Kazameini, C. Stachl, E. Cambria, & S. Eetemadi.** (2020). Bottom-up and top-down: Predicting personality with psycholinguistic and language model features. 2020 IEEE International Conference on Data Mining (ICDM),

**Minace, S., M. Minaei, & A. Abdolrashidi.** (2021). Deep-emotion: Facial expression recognition using attentional convolutional network. *Sensors*, 21(9), 3046.

**Mollahosseini, A., B. Hasani, & M. H. Mahoor.** (2017). Affectnet: A database for facial expression, valence, and arousal computing in the wild. *IEEE Transactions on Affective Computing*, 10(1), 18–31.

**Monti, F., F. Frasca, D. Eynard, D. Mannion, & M. Bronstein.** (2019). Fake news detection on social media

using geometric deep learning. arXiv preprint arXiv:1902.06673.

**Moritani, A., R. Ozaki, S. Sakamoto, H. Kameoka, & T. Taniguchi.** (2021). Stargan-based emotional voice conversion for japanese phrases. arXiv preprint arXiv:2104.01807.

**Murthy, A. R., & K. A. Kumar.** (2021). A Review of Different Approaches for Detecting Emotion from Text. *IOP Conference Series: Materials Science and Engineering*,

**Naeem, S., M. Iqbal, M. Saqib, M. Saad, M. S. Raza, Z. Ali, N. Akhtar, M. O. Beg, W. Shahzad, & M. U. Arshad.** (2020). Subspace gaussian mixture model for continuous urdu speech recognition using kaldi. 2020 14th International Conference on Open Source Systems and Technologies (ICOSST).

**Nandwani, P., & R. Verma.** (2021). A review on sentiment analysis and emotion detection from text. *Social Network Analysis and Mining*, 11(1), 1–19.

**Nguyen, D.-P., M.-C. Ho Ba Tho, & T.-T. Dao.** (2021). Enhanced facial expression recognition using 3D point sets and geometric deep learning. *Medical & Biological Engineering & Computing*, 59(6), 1235–1244.

**Olah, J., S. Baruah, D. Bose, & S. Narayanan.** (2021). Cross Domain Emotion Recognition using Few Shot Knowledge Transfer. arXiv preprint arXiv:2110.05021.

**Oloyede, M., G. Hancke, H. Myburgh, & A. Onumanyi.** (2019). A new evaluation function for face image enhancement in unconstrained environments using metaheuristic algorithms. *EURASIP Journal on Image and Video Processing*, 2019(1), 1–18.

**Oloyede, M. O., G. P. Hancke, & H. C. Myburgh.** (2018). Improving face recognition systems using a new image enhancement technique, hybrid features and the convolutional neural network. *IEEE Access*, 6, 75181–75191.

**Oloyede, M. O., G. P. Hancke, & H. C. Myburgh.** (2020). A review on face recognition systems: recent approaches and challenges. *Multimedia Tools and Applications*, 79(37), 27891–27922.

**Ribeiro, F. L., S. Bollmann, & A. M. Puckett.** (2021). Predicting the retinotopic organization of human visual cortex from anatomy using geometric deep learning. *NeuroImage*, 244, 118624.

**Rupapara, V., F. Rustam, H. F. Shahzad, A. Mehmood, I. Ashraf, & G. S. Choi.** (2021). Impact of SMOTE on imbalanced text features for toxic comments classification using RVVC model. *IEEE Access*, 9, 78621–78634.

**Sailunaz, K., M. Dhaliwal, J. Rokne, & R. Alhaji.** (2018). Emotion detection from text and speech: a survey. *Social Network Analysis and Mining*, 8(1), 1–26.

**Samara, A., L. Galway, R. Bond, & H. Wang.** (2019). Affective state detection via facial expression analysis within a human-computer interaction context. *Journal of Ambient Intelligence and Humanized Computing*, 10(6), 2175–2184.

- Shagdar, Z., M. Ullah, H. Ullah, & F. A. Cheikh.** (2021). Geometric deep learning for multi-object tracking: A brief review. 2021 9th European Workshop on Visual Information Processing (EUVIP),
- Singh, J., & R. Gill.** (2022). Multimodal Emotion Recognition System Using Machine Learning and Psychological Signals: A Review. *Soft Computing: Theories and Applications*, 657–666.
- Sreeja, P.S., & Mahalakshmi, G..** (2017). Emotion models: a review. *International Journal of Control Theory and Applications*, 10, 651–657.
- Susskind, J., G. Littlewort, M. Bartlett, J. Movellan, & A. Anderson.** (2007). Human and computer recognition of facial expressions of emotion. *Neuropsychologia*, 45(1), 152–162.
- Suzuki, T., A. Taya, & Y. Tobe.** (2021). VFep: 3D Graphic Face Representation Based on Voice-based Emotion Recognition. 2021 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops),
- Tariq, Z., S. K. Shah, & Y. Lee.** (2019). Speech emotion detection using iot based deep learning for health care. 2019 IEEE International Conference on Big Data (Big Data),
- Tarnowski, P., M. Kołodziej, A. Majkowski, & R. J. Rak.** (2017). Emotion recognition using facial expressions. *Procedia Computer Science*, 108, 1175–1184.
- Tomba, K., J. Dumoulin, E. Mugellini, O. Abou Khaled, & S. Hawila.** (2018). Stress Detection Through Speech Analysis. *ICETE* (1),
- Tripathi, S., A. Kumar, A. Ramesh, C. Singh, & P. Yenigalla.** (2019). Deep learning-based emotion recognition system using speech features and transcriptions. arXiv preprint arXiv:1906.05681.
- Tu, G., T. Xie, B. Liang, H. Wang, & R. Xu.** (2024). Adaptive Graph Learning for Multimodal Conversational Emotion Detection. *Proceedings of the AAAI Conference on Artificial Intelligence*,
- Tuhin, R. A., B. K. Paul, F. Nawrine, M. Akter, & A. K. Das.** (2019). An automated system of sentiment analysis from Bangla text using supervised learning techniques. 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS),
- Turabzadeh, S., H. Meng, R. M. Swash, M. Pleva, & J. Juhar.** (2018). Facial expression emotion detection for real-time embedded systems. *Technologies*, 6(1), 17.
- Verma, G., & H. Verma.** (2020). Hybrid-deep learning model for emotion recognition using facial expressions. *The Review of Socionetwork Strategies*, 14(2), 171–180.
- Yang, K., C. Wang, Z. Sarsenbayeva, B. Tag, T. Dingler, G. Wadley, & J. Goncalves.** (2021). Benchmarking commercial emotion detection systems using realistic distortions of facial image datasets. *The Visual Computer*, 37(6), 1447–1466.
- Zad, S., M. Heidari, H. James Jr, & O. Uzuner.** (2021). Emotion detection of textual data: An interdisciplinary survey. 2021 IEEE World AI IoT Congress (AIoT),
- Zamil, A. A. A., S. Hasan, S. M. J. Baki, J. M. Adam, & I. Zaman.** (2019). Emotion detection from speech signals using voting mechanism on classified frames. 2019 International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST),
- Zhang, D., L. Wu, C. Sun, S. Li, Q. Zhu, & G. Zhou.** (2019). Modeling both Context-and Speaker-Sensitive Dependence for Emotion Detection in Multi-speaker Conversations. *IJCAI*,
- Zhang, Y., J. Fu, D. She, Y. Zhang, S. Wang, & J. Yang.** (2018). Text Emotion Distribution Learning via Multi-Task Convolutional Neural Network. *IJCAI*,
- Zhao, Z., Q. Liu, & S. Wang.** (2021). Learning deep global multi-scale and local attention features for facial expression recognition in the wild. *IEEE Transactions on Image Processing*, 30, 6544–6556.
- Zhong, B., Z. Qin, S. Yang, J. Chen, N. Mudrick, M. Taub, R. Azevedo, & E. Lobaton.** (2017). Emotion recognition with facial expressions and physiological signals. 2017 IEEE symposium series on computational intelligence (SSCI),
- Zhong, P., D. Wang, & C. Miao.** (2019). Knowledge-enriched transformer for emotion detection in textual conversations. arXiv preprint arXiv:1909.10681.