

RESEARCH ARTICLE

Emotion-Aware Social Media Feed with Real- Time Emotion Recognition to Enhance Mental Health

Pinki Patra^{1,#}, Richa Sahu¹, Sonal Yadav¹, D Mohini¹, Chanchala Patel¹, Mahendra Kumar Soni¹

¹ Department of AI& AIML, Rungta College of Engineering & Technology, Bhilai, India-490024

*Corresponding Author: pinki.patra@rungta.ac.in

ABSTRACT:

Social media now plays a crucial role in everyday life by offering channels for exchanging information, communicating, and expressing oneself. However, users' mental health can be severely impacted by excessive and uncontrolled exposure to emotionally upsetting or negative content, which can result in social isolation, stress, anxiety, and despair. Conventional social media platforms frequently lack emotional intelligence and are ill-prepared to react to users' emotional states in a dynamic manner. In order to provide a customized and emotionally responsive user experience, this project suggests an Emotion-Aware Social Media Feed that makes use of real-time emotion recognition. The system can determine users' emotions while they engage with material by combining technologies including speech tone detection, natural language processing (NLP), and facial expression analysis. The platform prioritizes and curates information based on these emotional insights, limiting exposure to potentially dangerous content and encouraging postings that are uplifting, relaxing, or motivational when distressing indications are identified. Promoting emotional awareness, improving digital well-being, and offering early help or treatments when needed are the objectives. Furthermore, the system prioritizes ethical aspects such as fairness across various demographic groups, secure data handling, and informed user permission. In order to create better social media connections, this project intends to turn passive content consumption into an emotionally intelligent and mentally supporting experience by fusing artificial intelligence with affective computing concepts

KEYWORDS: Emotion recognition, mental health, affective computing, personalized social media, digital well-being.

1. Introduction

Social media platforms have become ubiquitous, deeply embedding themselves into the fabric of daily human interaction. However, a growing body of psychological and sociological research indicates a strong correlation between excessive social media use and adverse mental health outcomes, including increased rates of anxiety, depression, social isolation, and stress.

Affective computing is the study and development of systems and devices that can recognize, interpret, process, and simulate human affects. It is an interdisciplinary field spanning computer science, psychology, and cognitive science. While some core ideas in the field may be traced as far back as to early philosophical inquiries into emotion, the modern idea originated with Rosalind Picard's 1995 paper entitled "Affective Computing"

and her 1997 book of the same name published by MIT Press. One motivation for researching affective computing is the ability to give machines emotional intelligence, including simulating empathy. The goal is that a machine should interpret the emotional state of humans and adapt its behavior to those emotions, responding appropriately. Recent experimental research has shown that subtle affective haptic feedback can shape human reward learning and mobile interaction behavior, suggesting that affective computing systems may not only interpret emotional states but also actively modulate user actions through emotion-laden outputs [1].

Detecting emotional information usually begins with passive sensors that capture data about the user's physical state or behavior without interpreting the input. The data gathered is analogous to the cues humans use to perceive emotions in others. For example, a video camera might capture facial expres-

sions, body posture, and gestures, while a microphone might capture speech. Other sensors detect emotional cues by directly measuring physiological data, such as skin temperature and galvanic resistance [2].

Recognizing emotional information requires the extraction of meaningful patterns from the gathered data. This is done using machine learning techniques that process different modalities, such as speech recognition, natural language processing, or facial expression detection. The goal of most of these techniques is to produce labels that would match the labels a human would give in the same situation. For example, if a person makes a facial expression furrowing their brow, then the computer vision system might be trained to label their face as appearing "confused" or as "concentrating" or "slightly negative" (as opposed to positive, which it might say if they were smiling in a happy-appearing way). This response is based on the data used to train the system. These labels may or may not correspond to what the person is actually feeling [3].

Another area within affective computing is the design of computational devices proposed to exhibit either innate emotional capabilities or that are capable of convincingly simulating emotions. A more practical approach, based on current technological capabilities, is the simulation of emotions in conversational agents in order to enrich and facilitate interactivity between human and machine [4].

Marvin Minsky, one of the pioneering computer scientists in artificial intelligence, relates emotions to the broader issues of machine intelligence stating in *The Emotion Machine* that emotion is "not especially different from the processes that we call 'thinking.'" The innovative approach "digital humans" or virtual humans includes an attempt to give these programs, which simulate humans, an emotional dimension as well, including reactions, facial expressions, and gestures in accordance with the reaction that a real person would have in certain emotionally stimulating situations [5].

Emotion in machines often refers to emotion in computational, often AI-based, systems. As a result, the terms 'emotional AI' is being used. Some modern large language models simulate emotions in their chats with humans. ChatGPT's simulated emotion leans more positive than that of most human responses [6].

In psychology, cognitive science, and neuroscience, there have been two main approaches for describing how humans perceive and classify emotion: continuous or categorical. The continuous approach tends to use dimensions such as negative vs. positive, calm vs. aroused [7].

The categorical approach tends to use discrete classes such as happy, sad, angry, fearful, surprise, and disgust. Different kinds of machine learning regression and classification models are used for machines to produce continuous or discrete labels. Sometimes, models are also built that allow combinations across the categories (e.g. a happy-surprised face or a fearful-surprised face) [8].

Various changes in the autonomic nervous system can in-

directly alter a person's speech, and affective technologies can leverage this information to recognize emotion. For example, speech produced in a state of fear, anger, or joy becomes fast, loud, and precisely enunciated, with a higher and wider range in pitch, whereas emotions such as tiredness, boredom, or sadness tend to generate slow, low-pitched, and slurred speech. Some emotions have been found to be more easily computationally identified, such as anger or approval [9].

Emotional speech processing technologies recognize the user's emotional state using computational analysis of speech features. Vocal parameters and prosodic features such as pitch variables and speech rate can be analyzed through pattern recognition techniques [10].

Speech analysis is an effective method of identifying affective state, having an average reported accuracy of 70-80% in research from 2003 and 2006. These systems tend to outperform average human accuracy (approximately 60%) but are less accurate than systems which employ other modalities for emotion detection, such as physiological states or facial expressions. However, since many speech characteristics are independent of semantics or culture, this technique is considered to be a promising route for further research [11].

The process of speech/text affect detection requires the creation of a reliable database, knowledge base, or vector space model, broad enough to fit every need for its application, as well as the selection of a successful classifier which will allow for quick and accurate emotion identification [12].

As of 2010, the most frequently used classifiers were linear discriminant classifiers (LDC), k-nearest neighbor (k-NN), Gaussian mixture model (GMM), support vector machines (SVM), artificial neural networks (ANN), decision tree algorithms, and hidden Markov models (HMMs). Various studies showed that choosing the appropriate classifier can significantly enhance the overall performance of the system. The list below gives a brief description of each algorithm: [13]

LDC - Classification happens based on the value obtained from the linear combination of the feature values, which are usually provided in the form of vector features [14].

k-NN - Classification happens by locating the object in the feature space, and comparing it with the k nearest neighbors (training examples). The majority vote decides on the classification [15].

GMM - A probabilistic model used for representing the existence of subpopulations within the overall population. Each sub-population is described using the mixture distribution, which allows for classification of observations into the subpopulations [16].

Decision tree algorithms - work based on following a decision tree in which leaves represent the classification outcome, and branches represent the conjunction of subsequent features that lead to the classification [17].

HMMs - a statistical Markov model in which the states and state transitions are not directly available to observation. Instead, the series of outputs dependent on the states are visi-

ble. In the case of affect recognition, the outputs represent the sequence of speech feature vectors, which allow the deduction of states' sequences through which the model progressed. The states can consist of various intermediate steps in the expression of an emotion, and each of them has a probability distribution over the possible output vectors. The states' sequences allow us to predict the affective state which we are trying to classify, and this is one of the most commonly used techniques within the area of speech affect detection [18].

2. Related Work

The intersection of affective computing and social media is a rapidly expanding field. Early research laid the foundational framework for affective computing. In recent years, researchers have utilized social media text to retrospectively analyze public sentiment and predict mental health trends.

It has been proven that with enough acoustic evidence available, the emotional state of a person can be classified by a set of majority voting classifiers. The proposed set of classifiers is based on three main classifiers: kNN, C4.5 and SVM-RBF Kernel. This set achieves better performance than each basic classifier taken separately. It is compared with two other sets of classifiers: one-against-all (OAA) multiclass SVM with Hybrid kernels and the set of classifiers which consists of the following two basic classifiers: C5.0 and Neural Network. The proposed variant achieves better performance than the other two sets of classifiers [19].

The vast majority of present systems are data-dependent. This creates one of the biggest challenges in detecting emotions based on speech, as it implicates choosing an appropriate database to train the classifier. Most of the current available data was obtained from actors, and is thus a representation of archetypal emotions. Those so-called "acted databases" are usually based on the Basic Emotions theory by Paul Ekman, which assumes the existence of six basic emotions (anger, fear, disgust, surprise, joy, sadness), the others a mix of those six. Nevertheless, these still offer high audio quality and balanced classes (although often too few), which contribute to high success rates in recognizing emotions [20].

However, for real-life application, naturalistic data is preferred. A naturalistic database can be produced by observation and analysis of subjects in their natural context. Ultimately, such a database should allow the system to recognize emotions based on their context as well as work out the goals and outcomes of the interaction. The nature of this type of data allows for authentic real-life implementation, due to the fact that it describes states naturally occurring during human-computer interaction (HCI) [1].

Despite the numerous advantages which naturalistic data has over acted data, it is difficult to obtain and usually has low emotional intensity. Moreover, data obtained in a natural context has lower signal quality, due to surroundings, noise, and the subjects' distance from the microphone. The first attempt

to produce such a database was the FAU Aibo Emotion Corpus for CEICES (Combining Efforts for Improving Automatic Classification of Emotional User States), which was developed based on a realistic context of children (age 10-13) playing with Sony's Aibo robot pet. Likewise, producing one standard database for all emotional research would provide a method of evaluating and comparing different affect recognition systems [2].

The complexity of the affect recognition process increases with the number of classes (affects) and speech descriptors used within the classifier. It is, therefore, crucial to select only the most relevant features in order to assure the ability of the model to successfully identify emotions and increase performance, which is particularly significant to real-time detection. The range of possible choices is vast, with some studies mentioning the use of over 200 distinct features. It is crucial to identify those that are redundant and undesirable in order to optimize the system and increase the success rate of correct emotion detection. The most common speech characteristics are categorized into the following groups [3].

The detection and processing of facial expression is achieved through various methods such as optical flow, hidden Markov models, neural network processing, or active appearance models. More than one modality can be combined or fused (multimodal recognition, e.g. facial expressions and speech prosody, facial expressions and hand gestures, or facial expressions with speech and text for multimodal data and metadata analysis) to provide a more robust estimation of the subject's emotional state [4].

Creation of an emotion database is a difficult and time-consuming task. However, database creation is an essential step in the formation of a system that will recognize human emotions. Most publicly available emotion databases include posed facial expressions only. In posed expression databases, the participants are asked to display different basic emotional expressions. In the alternative spontaneous expression databases, the expressions are natural. Spontaneous emotion elicitation requires significant effort in the selection of proper stimuli which can lead to a rich display of intended emotions. Secondly, the process requires emotions to be tagged by trained individuals manually, which makes the databases highly reliable. Since perception of expressions and their intensity is subjective in nature, the annotation by experts is essential for the purpose of validation [5].

Researchers work with three types of databases: a database of peak expression images only, a database of image sequences portraying an emotion from neutral to its peak, and video clips with emotional annotations. Many facial expression databases have been created and made public for expression recognition purposes. Two widely used databases are CK+ and JAFFE [6].

During the late 1960's, Paul Ekman proposed the idea that facial expressions of emotion are not culturally determined, but universal. He discovered this by doing cross-cultural research in Papua, New Guinea on the Fore Tribesmen. Thus, he

suggested that they are biological in origin and can, therefore, be safely and correctly categorized [7].

However, in the 1990s, Ekman expanded his list of basic emotions, including a range of positive and negative emotions, not all of which are encoded in facial muscles. The new emotions are: [8]

A system has been conceived by psychologists in order to formally categorize the physical expression of emotions on faces. The central concept of the Facial Action Coding System, or FACS, as created by Paul Ekman and Wallace V. Friesen in 1978 based on earlier work by Carl-Herman Hjortsj. Action units (AU) are, basically, a contraction or a relaxation of one or more muscles. Psychologists have proposed the following classification of six basic emotions according to their action units ("+" here means "and"): [9]

As with every computational practice, some obstacles need to be surpassed in order to fully unlock the hidden potential of in affect detection by facial processing. In the early days of AI-based detection (speech recognition, face recognition, and affect recognition), the accuracy of modeling and tracking was an issue. As hardware evolves, this lack of accuracy fades. This, though, leaves behind noise issues. However, methods for noise removal exist, including neighborhood averaging, linear Gaussian smoothing, median filtering, or newer methods such as the Bacterial Foraging Optimization Algorithm [10].

The lack of rotational movement freedom. Affect detection works very well with frontal use, but upon rotating the head more than 20 degrees, "[there have] been problems" [11].

Facial expressions do not always correspond to an underlying emotion that matches them (e.g. they can be posed or faked, or a person can feel emotions but maintain a "poker face") [12].

FACS did not include dynamics, which can help disambiguate (e.g. smiles of genuine happiness tend to have different dynamics than "try to look happy" smiles.) [13]

The FACS combinations do not correspond in a 1:1 way with the emotions that the psychologists originally proposed (note that this lack of a 1:1 mapping also occurs in speech recognition with homophones and homonyms and many other sources of ambiguity, and may be mitigated by bringing in other channels of information) [14].

Gestures can be used as a means of detecting the particular emotional state of a user, especially when used in conjunction with speech and face recognition. Gestures could be simple reflexive responses, like lifting one's shoulders when one does not know the answer to a question, or they could be complex and meaningful, as when communicating with sign language. Without making use of any object or surrounding environment, we can wave our hands, clap or beckon. On the other hand, when using objects, we can point at them, move, touch, or handle them. A computer should be able to recognize these gestures, analyze the context, and respond in a meaningful way in order to be used for Human-Computer Interaction [15].

Gestures can be detected using many methods. There are

two main approaches towards gesture recognition: 3D model based and appearance-based. The foremost method makes use of 3D information of key elements of the body parts in order to obtain several important parameters, like palm position or joint angles. On the other hand, appearance-based systems use images or videos to for direct interpretation. Hand gestures have been a common focus of body gesture detection methods [16].

This could be used to detect a user's affective state by monitoring and analyzing their physiological signs. These signs range from changes in heart rate and skin conductance to minute contractions of the facial muscles and changes in facial blood flow. This area is gaining momentum and we are now seeing real products that implement the techniques. The four main physiological signs that are usually analyzed are blood volume pulse, galvanic skin response, facial electromyography, and facial color patterns. For instance, these methods have been used to examine for symptoms of depression [17].

A subject's blood volume pulse (BVP) can be measured by a process called photoplethysmography, which produces a graph indicating blood flow through the extremities. The peaks of the waves indicate a cardiac cycle where the heart has pumped blood to the extremities. If the subject experiences fear or is startled, their heart usually 'jumps' and beats quickly for some time, causing the amplitude of the cardiac cycle to increase. This can clearly be seen on a photoplethysmograph when the distance between the trough and the peak of the wave has decreased. As the subject calms down, and as the body's inner core expands, allowing more blood to flow back to the extremities, the cycle will return to normal [18].

Researchers shone infrared light on the skin using special sensor hardware, and the amount of light reflected was measured. The amount of reflected and transmitted light correlates to the BVP as light is absorbed by hemoglobin which is found richly in the bloodstream [19].

It can be cumbersome to ensure that the sensor shining an infrared light and monitoring the reflected light is always pointing at the same extremity, especially because subjects often stretch and readjust their position while using a computer [20].

There are other factors that can affect one's blood volume pulse. Since it is a measure of blood flow through the extremities, if the subject feels particularly hot or cold, then their body may allow more or less blood to flow to the extremities regardless of the subject's emotional state [1].

Facial electromyography is a technique used to measure the electrical activity of the facial muscles by amplifying the tiny electrical impulses that are generated by muscle fibers when they contract [2].

The corrugator supercilii muscle, also known as the 'frowning' muscle, draws the brow down into a frown. Therefore, it is the best test for negative, unpleasant emotional responses. The zygomaticus major muscle is responsible for pulling the corners of the mouth back when one smiles, and therefore is the muscle used to test for a positive emotional

response [3].

Galvanic skin response (GSR) is an outdated term for a more general phenomenon known as Electrodermal activity, or EDA. EDA is a general phenomena whereby the skin's electrical properties change. The skin is innervated by the sympathetic nervous system, so measuring its resistance or conductance provides a way to quantify small changes in the sympathetic branch of the autonomic nervous system. As sweat glands are activated, even before the skin feels sweaty, the level of the EDA can be captured (usually using conductance) and used to discern small changes in autonomic arousal. The more aroused a subject is, the greater the skin conductance tends to be [4].

Skin conductance is often measured using two small silver-silver chloride electrodes placed somewhere on the skin and applying a small voltage between them. To maximize comfort and reduce irritation the electrodes can be placed on the wrist, legs, or feet, which leaves the hands fully free for daily activity [5].

The surface of the human face is innervated with a large network of blood vessels. Blood flow variations in these vessels yield visible color changes on the face. Whether or not facial emotions activate facial muscles, variations in blood flow, blood pressure, glucose levels, and other changes occur. Also, the facial color signal is independent from that provided by facial muscle movements [6].

Approaches are based on facial color changes. Delaunay triangulation is used to create the triangular local areas. Some of these triangles which define the interior of the mouth and eyes (sclera and iris) are removed. The left triangular areas' pixels are used to create feature vectors. Evidence shows that converting the pixel color of the standard RGB color space to a color space-such as oRGB color space or LMS channels-allows for better performance when dealing with faces. So, vectors are mapped onto the better color space and decomposed into red-green and yellow-blue channels. Then deep learning methods find equivalent emotions [7].

Aesthetics, in the world of art and photography, refer to the principles of nature and appreciation of beauty. Judging beauty and other aesthetic qualities is a highly subjective task. Computer scientists at Penn State treat the challenge of automatically inferring the aesthetic quality of pictures using their visual content as a machine learning problem, with a peer-rated on-line photo sharing website as a data source. They extract certain visual features based on the intuition that they can discriminate between aesthetically pleasing and displeasing images [8].

Affection influences learning state. Using affective computing technology, computers can judge the learners' affection and learning state by recognizing their facial expressions. In education, a teacher can use this analysis result to understand the student's learning and accepting ability, and then formulate reasonable teaching plans. At the same time, they can pay attention to students' inner feelings, which is helpful to students'

psychological health. Especially in asynchronous education, due to the separation of time and space, there is no emotional incentive between teachers and students for two-way communication. Without the atmosphere brought by traditional classroom learning, students are easily bored, which can affect learning outcomes. Applying affective computing in distance education system can improve this situation. Emotional AI can provide students with AI-based learning support, which benefits students' cognitive and emotional outcomes [9].

The applications of sensory computing may contribute to improving road safety. For example, a car can monitor the emotion of all occupants and engage in additional safety measures, such as alerting other vehicles if it detects the driver to be angry. In addition, affective computing systems may allow various interventions such as driver assistance systems adjusted according to the driver's stress level as well as minimal and direct interventions to change the emotional state of the driver [10].

Social robots, as well as a growing number of robots used in healthcare, benefit from emotional awareness because they can better judge users' and patients' emotional states, altering their actions/programming appropriately. This is especially important in countries with growing aging populations and a lack of younger workers to address their needs [11].

Affective computing is also being applied to the development of communicative technologies for use by people with autism. The affective component of a text is also increasingly gaining attention, particularly its role in the so-called emotional or emotive Internet [12].

Affective video games can access their players' emotional states through biofeedback devices. A particularly simple form of biofeedback is a gamepad that measures the pressure with which a button is pressed: this has been shown to correlate strongly with the players' level of arousal. At the other end of the scale are brain-computer interfaces. Affective games have been used in medical research to support the emotional development of autistic children [13].

Training methods of psychomotor operations, such as steering and maneuvering, are used in various fields such as aviation, transportation and medicine. Integrating affective computing capabilities in these types of training systems, in accordance with the adaptive automation approach, has been found to be effective in improving the quality of training and shortening the required training duration [14].

Affective computing has potential applications in human-computer interaction, such as affective mirrors that allow the user to see how they perform, emotion monitoring agents that send a warning before one sends an angry email, or even music players that can select tracks based on mood [15].

One idea put forth by the Romanian researcher Dr. Nicu Sebe in an interview is the analysis of a person's face while they are using a certain product (he mentioned ice cream as an example). Companies would then be able to use such analysis to infer whether their product will or will not be well received

by the respective market [16].

One could also use affective state recognition in order to judge the impact of a TV advertisement through a real-time video recording of that person and through the subsequent study of their facial expression. Averaging the results obtained on a large group of subjects, one can tell whether that commercial (or movie) has the desired effect and what the elements which interest the watcher most are [17].

Within the field of human-computer interaction, Rosalind Picard's cognitivist or "information model" concept of emotion has been criticized by and contrasted with the "post-cognitivist" or "interactional" pragmatist approach taken by Kirsten Boehner and others which views emotion as inherently social [18].

Picard's focus is human-computer interaction, and her goal for affective computing is to "give computers the ability to recognize, express, and in some cases, 'have' emotions". In contrast, the interactional approach seeks to help "people to understand and experience their own emotions" and to improve computer-mediated interpersonal communication. It does not necessarily seek to map emotion into an objective mathematical model for machine interpretation, but rather let humans make sense of each other's emotional expressions in open-ended ways that might be ambiguous, subjective, and sensitive to context [19].

Picard's critics describe her concept of emotion as "objective, internal, private, and mechanistic". They say it reduces emotion to a discrete psychological signal occurring inside the body that can be measured and which is an input to cognition, undercutting the complexity of emotional experience [20].

The interactional approach asserts that though emotion has biophysical aspects, it is "culturally grounded, dynamically experienced, and to some degree constructed in action and interaction". Put another way, it considers "emotion as a social and cultural product experienced through our interactions" [1].

Usage of affective computing tools such as ChatGPT has led to some individuals developing parasocial relationships with the LLMs. One woman who was interviewed by The New York Times chatted with her "A.I. boyfriend", ChatGPT, for up to fifty-six hours every week. These kinds of interactions with affective computing tools may increase feelings of depression or loneliness in those with underlying mental health conditions, resulting in a phenomenon often coined "Chatbot psychosis". Matthew Raine and Megan Garcia, who lost their sons in A.I.-related incidents, have sued OpenAI for encouraging these kinds of relationships with ChatGPT [2].

Although affective computing systems have many potential uses, some of those uses pose ethical concern in the scientific community. Usage of affective computing systems could analyze people's facial expressions in public without explicit consent from individuals. There is also a risk of affective computing systems being used to manipulate an audience's emotions. Additionally, systems that help others may pose small trade-off risks (such as arousing feelings that are not connected

to the real world). Researchers are also concerned about potential psychological effects of affective computing that is used in a sexual context (ex. robotic sex dolls), though no evidence exists yet to prove these claims. Users of affective computing systems trust them more when there is an ethical contract between the system and the user [3].

Hudlicka, Eva (2003). "To feel or not to feel: The role of affect in human-computer interaction". *International Journal of Human-Computer Studies*. 59 (1-2): 1-32. CiteSeerX 10.1.1.180.6429. doi:10.1016/s1071-5819(03)00047-8 [4].

Scherer, Klaus R; Bnziger, Tanja; Roesch, Etienne B (2010). *A Blueprint for Affective Computing: A Sourcebook and Manual*. Oxford: Oxford University Press. ISBN 978-0-19-956670-9 [5].

Affective computing is the study and development of systems and devices that can recognize, interpret, process, and simulate human affects. It is an interdisciplinary field spanning computer science, psychology, and cognitive science. While some core ideas in the field may be traced as far back as to early philosophical inquiries into emotion, the modern idea originated with Rosalind Picard's 1995 paper entitled "Affective Computing" and her 1997 book of the same name published by MIT Press. One motivation for researching affective computing is the ability to give machines emotional intelligence, including simulating empathy. The goal is that a machine should interpret the emotional state of humans and adapt its behavior to those emotions, responding appropriately. Recent experimental research has shown that subtle affective haptic feedback can shape human reward learning and mobile interaction behavior, suggesting that affective computing systems may not only interpret emotional states but also actively modulate user actions through emotion-laden outputs [6].

Detecting emotional information usually begins with passive sensors that capture data about the user's physical state or behavior without interpreting the input. The data gathered is analogous to the cues humans use to perceive emotions in others. For example, a video camera might capture facial expressions, body posture, and gestures, while a microphone might capture speech. Other sensors detect emotional cues by directly measuring physiological data, such as skin temperature and galvanic resistance [7].

Recognizing emotional information requires the extraction of meaningful patterns from the gathered data. This is done using machine learning techniques that process different modalities, such as speech recognition, natural language processing, or facial expression detection. The goal of most of these techniques is to produce labels that would match the labels a human would give in the same situation. For example, if a person makes a facial expression furrowing their brow, then the computer vision system might be trained to label their face as appearing "confused" or as "concentrating" or "slightly negative" (as opposed to positive, which it might say if they were smiling in a happy-appearing way). This response is based on the data used to train the system. These labels may or may not

correspond to what the person is actually feeling [8].

Another area within affective computing is the design of computational devices proposed to exhibit either innate emotional capabilities or that are capable of convincingly simulating emotions. A more practical approach, based on current technological capabilities, is the simulation of emotions in conversational agents in order to enrich and facilitate interactivity between human and machine [9].

Marvin Minsky, one of the pioneering computer scientists in artificial intelligence, relates emotions to the broader issues of machine intelligence stating in The Emotion Machine that emotion is "not especially different from the processes that we call 'thinking.'" The innovative approach "digital humans" or virtual humans includes an attempt to give these programs, which simulate humans, an emotional dimension as well, including reactions, facial expressions, and gestures in accordance with the reaction that a real person would have in certain emotionally stimulating situations [10].

Emotion in machines often refers to emotion in computational, often AI-based, systems. As a result, the terms 'emotional AI' is being used. Some modern large language models simulate emotions in their chats with humans. ChatGPT's simulated emotion leans more positive than that of most human responses [11].

In psychology, cognitive science, and neuroscience, there have been two main approaches for describing how humans perceive and classify emotion: continuous or categorical. The continuous approach tends to use dimensions such as negative vs. positive, calm vs. aroused [12].

3. Methodology

The Emotion-Aware Feed operates through a continuous, closed-loop process of emotion detection, algorithmic curation, and user feedback. The emotion recognition module utilizes three distinct data streams to ensure high accuracy and robustness against individual behavioral variations.

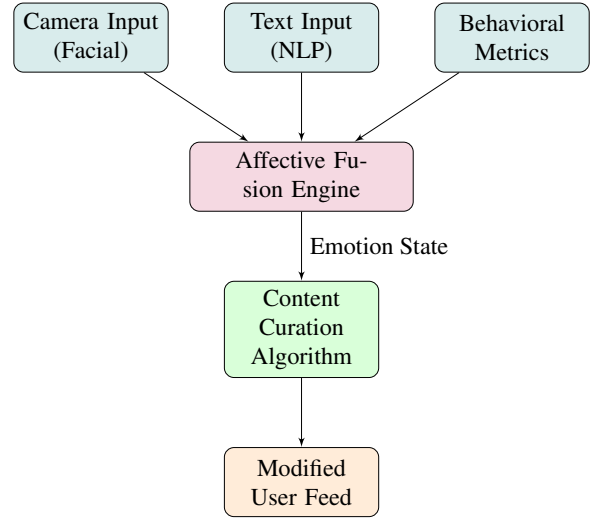


Figure 1: Architecture of the Emotion-Aware Feed System

Algorithm 1 Dynamic Feed Adjustment based on ESI

Require: User ESI $E \in [-1, 1]$, Content Pool C , Content Valence V_c

Ensure: Sorted Feed F

```

1: for each post  $c \in C$  do
2:    $BaseScore \leftarrow CalculateRelevance(c)$ 
3:   if  $E < -0.5$  then ▷ User is distressed
4:     if  $V_c < 0$  then ▷ Content is negative
5:        $c.Score \leftarrow BaseScore \times 0.1$  ▷ Suppress
6:     else if  $V_c > 0.7$  then ▷ Content is highly positive
7:        $c.Score \leftarrow BaseScore \times 1.5$  ▷ Boost
8:     end if
9:   else
10:     $c.Score \leftarrow BaseScore$  ▷ Normal feed
11:   end if
12: end for
13:  $F \leftarrow SortDescending(C \text{ by } Score)$ 
14: return  $F$ 

```

The categorical approach tends to use discrete classes such as happy, sad, angry, fearful, surprise, and disgust. Different kinds of machine learning regression and classification models are used for machines to produce continuous or discrete labels. Sometimes, models are also built that allow combinations across the categories (e.g. a happy-surprised face or a fearful-surprised face) [13].

Various changes in the autonomic nervous system can indirectly alter a person's speech, and affective technologies can leverage this information to recognize emotion. For example, speech produced in a state of fear, anger, or joy becomes fast, loud, and precisely enunciated, with a higher and wider range in pitch, whereas emotions such as tiredness, boredom, or sad-

ness tend to generate slow, low-pitched, and slurred speech. Some emotions have been found to be more easily computationally identified, such as anger or approval [14].

Emotional speech processing technologies recognize the user's emotional state using computational analysis of speech features. Vocal parameters and prosodic features such as pitch variables and speech rate can be analyzed through pattern recognition techniques [15].

Speech analysis is an effective method of identifying affective state, having an average reported accuracy of 70-80% in research from 2003 and 2006. These systems tend to outperform average human accuracy (approximately 60%) but are less accurate than systems which employ other modalities for emotion detection, such as physiological states or facial expressions. However, since many speech characteristics are independent of semantics or culture, this technique is considered to be a promising route for further research [16].

The process of speech/text affect detection requires the creation of a reliable database, knowledge base, or vector space model, broad enough to fit every need for its application, as well as the selection of a successful classifier which will allow for quick and accurate emotion identification [17].

As of 2010, the most frequently used classifiers were linear discriminant classifiers (LDC), k-nearest neighbor (k-NN), Gaussian mixture model (GMM), support vector machines (SVM), artificial neural networks (ANN), decision tree algorithms, and hidden Markov models (HMMs). Various studies showed that choosing the appropriate classifier can significantly enhance the overall performance of the system. The list below gives a brief description of each algorithm: [18]

LDC - Classification happens based on the value obtained from the linear combination of the feature values, which are usually provided in the form of vector features [19].

k-NN - Classification happens by locating the object in the feature space, and comparing it with the k nearest neighbors (training examples). The majority vote decides on the classification [20].

GMM - A probabilistic model used for representing the existence of subpopulations within the overall population. Each sub-population is described using the mixture distribution, which allows for classification of observations into the subpopulations [1].

Decision tree algorithms - work based on following a decision tree in which leaves represent the classification outcome, and branches represent the conjunction of subsequent features that lead to the classification [2].

HMMs - a statistical Markov model in which the states and state transitions are not directly available to observation. Instead, the series of outputs dependent on the states are visible. In the case of affect recognition, the outputs represent the sequence of speech feature vectors, which allow the deduction of states' sequences through which the model progressed. The states can consist of various intermediate steps in the expression of an emotion, and each of them has a probability

distribution over the possible output vectors. The states' sequences allow us to predict the affective state which we are trying to classify, and this is one of the most commonly used techniques within the area of speech affect detection [3].

It has been proven that with enough acoustic evidence available, the emotional state of a person can be classified by a set of majority voting classifiers. The proposed set of classifiers is based on three main classifiers: kNN, C4.5 and SVM-RBF Kernel. This set achieves better performance than each basic classifier taken separately. It is compared with two other sets of classifiers: one-against-all (OAA) multiclass SVM with Hybrid kernels and the set of classifiers which consists of the following two basic classifiers: C5.0 and Neural Network. The proposed variant achieves better performance than the other two sets of classifiers [4].

The vast majority of present systems are data-dependent. This creates one of the biggest challenges in detecting emotions based on speech, as it implicates choosing an appropriate database to train the classifier. Most of the current available data was obtained from actors, and is thus a representation of archetypal emotions. Those so-called "acted databases" are usually based on the Basic Emotions theory by Paul Ekman, which assumes the existence of six basic emotions (anger, fear, disgust, surprise, joy, sadness), the others a mix of those six. Nevertheless, these still offer high audio quality and balanced classes (although often too few), which contribute to high success rates in recognizing emotions [5].

However, for real-life application, naturalistic data is preferred. A naturalistic database can be produced by observation and analysis of subjects in their natural context. Ultimately, such a database should allow the system to recognize emotions based on their context as well as work out the goals and outcomes of the interaction. The nature of this type of data allows for authentic real-life implementation, due to the fact that it describes states naturally occurring during human-computer interaction (HCI) [6].

Despite the numerous advantages which naturalistic data has over acted data, it is difficult to obtain and usually has low emotional intensity. Moreover, data obtained in a natural context has lower signal quality, due to surroundings, noise, and the subjects' distance from the microphone. The first attempt to produce such a database was the FAU Aibo Emotion Corpus for CEICES (Combining Efforts for Improving Automatic Classification of Emotional User States), which was developed based on a realistic context of children (age 10-13) playing with Sony's Aibo robot pet. Likewise, producing one standard database for all emotional research would provide a method of evaluating and comparing different affect recognition systems [7].

The complexity of the affect recognition process increases with the number of classes (affects) and speech descriptors used within the classifier. It is, therefore, crucial to select only the most relevant features in order to assure the ability of the model to successfully identify emotions and increase perfor-

mance, which is particularly significant to real-time detection. The range of possible choices is vast, with some studies mentioning the use of over 200 distinct features. It is crucial to identify those that are redundant and undesirable in order to optimize the system and increase the success rate of correct emotion detection. The most common speech characteristics are categorized into the following groups [8].

The detection and processing of facial expression is achieved through various methods such as optical flow, hidden Markov models, neural network processing, or active appearance models. More than one modality can be combined or fused (multimodal recognition, e.g. facial expressions and speech prosody, facial expressions and hand gestures, or facial expressions with speech and text for multimodal data and metadata analysis) to provide a more robust estimation of the subject's emotional state [9].

Creation of an emotion database is a difficult and time-consuming task. However, database creation is an essential step in the formation of a system that will recognize human emotions. Most publicly available emotion databases include posed facial expressions only. In posed expression databases, the participants are asked to display different basic emotional expressions. In the alternative spontaneous expression databases, the expressions are natural. Spontaneous emotion elicitation requires significant effort in the selection of proper stimuli which can lead to a rich display of intended emotions. Secondly, the process requires emotions to be tagged by trained individuals manually, which makes the databases highly reliable. Since perception of expressions and their intensity is subjective in nature, the annotation by experts is essential for the purpose of validation [10].

Researchers work with three types of databases: a database of peak expression images only, a database of image sequences portraying an emotion from neutral to its peak, and video clips with emotional annotations. Many facial expression databases have been created and made public for expression recognition purposes. Two widely used databases are CK+ and JAFFE [11].

During the late 1960's, Paul Ekman proposed the idea that facial expressions of emotion are not culturally determined, but universal. He discovered this by doing cross-cultural research in Papua, New Guinea on the Fore Tribesmen. Thus, he suggested that they are biological in origin and can, therefore, be safely and correctly categorized [12].

However, in the 1990s, Ekman expanded his list of basic emotions, including a range of positive and negative emotions, not all of which are encoded in facial muscles. The new emotions are: [13]

A system has been conceived by psychologists in order to formally categorize the physical expression of emotions on faces. The central concept of the Facial Action Coding System, or FACS, as created by Paul Ekman and Wallace V. Friesen in 1978 based on earlier work by Carl-Herman Hjortsj. Action units (AU) are, basically, a contraction or a relaxation

of one or more muscles. Psychologists have proposed the following classification of six basic emotions according to their action units ("+" here means "and"): [14]

As with every computational practice, some obstacles need to be surpassed in order to fully unlock the hidden potential of in affect detection by facial processing. In the early days of AI-based detection (speech recognition, face recognition, and affect recognition), the accuracy of modeling and tracking was an issue. As hardware evolves, this lack of accuracy fades. This, though, leaves behind noise issues. However, methods for noise removal exist, including neighborhood averaging, linear Gaussian smoothing, median filtering, or newer methods such as the Bacterial Foraging Optimization Algorithm [15].

The lack of rotational movement freedom. Affect detection works very well with frontal use, but upon rotating the head more than 20 degrees, "[there have] been problems" [16].

Facial expressions do not always correspond to an underlying emotion that matches them (e.g. they can be posed or faked, or a person can feel emotions but maintain a "poker face") [17].

FACS did not include dynamics, which can help disambiguate (e.g. smiles of genuine happiness tend to have different dynamics than "try to look happy" smiles.) [18]

The FACS combinations do not correspond in a 1:1 way with the emotions that the psychologists originally proposed (note that this lack of a 1:1 mapping also occurs in speech recognition with homophones and homonyms and many other sources of ambiguity, and may be mitigated by bringing in other channels of information) [19].

Gestures can be used as a means of detecting the particular emotional state of a user, especially when used in conjunction with speech and face recognition. Gestures could be simple reflexive responses, like lifting one's shoulders when one does not know the answer to a question, or they could be complex and meaningful, as when communicating with sign language. Without making use of any object or surrounding environment, we can wave our hands, clap or beckon. On the other hand, when using objects, we can point at them, move, touch, or handle them. A computer should be able to recognize these gestures, analyze the context, and respond in a meaningful way in order to be used for Human-Computer Interaction [20].

Gestures can be detected using many methods. There are two main approaches towards gesture recognition: 3D model based and appearance-based. The foremost method makes use of 3D information of key elements of the body parts in order to obtain several important parameters, like palm position or joint angles. On the other hand, appearance-based systems use images or videos to for direct interpretation. Hand gestures have been a common focus of body gesture detection methods [1].

This could be used to detect a user's affective state by monitoring and analyzing their physiological signs. These signs range from changes in heart rate and skin conductance to minute contractions of the facial muscles and changes in facial blood flow. This area is gaining momentum and we are now

seeing real products that implement the techniques. The four main physiological signs that are usually analyzed are blood volume pulse, galvanic skin response, facial electromyography, and facial color patterns. For instance, these methods have been used to examine for symptoms of depression [2].

A subject's blood volume pulse (BVP) can be measured by a process called photoplethysmography, which produces a graph indicating blood flow through the extremities. The peaks of the waves indicate a cardiac cycle where the heart has pumped blood to the extremities. If the subject experiences fear or is startled, their heart usually 'jumps' and beats quickly for some time, causing the amplitude of the cardiac cycle to increase. This can clearly be seen on a photoplethysmograph when the distance between the trough and the peak of the wave has decreased. As the subject calms down, and as the body's inner core expands, allowing more blood to flow back to the extremities, the cycle will return to normal [3].

Researchers shone infrared light on the skin using special sensor hardware, and the amount of light reflected was measured. The amount of reflected and transmitted light correlates to the BVP as light is absorbed by hemoglobin which is found richly in the bloodstream [4].

It can be cumbersome to ensure that the sensor shining an infrared light and monitoring the reflected light is always pointing at the same extremity, especially because subjects often stretch and readjust their position while using a computer [5].

There are other factors that can affect one's blood volume pulse. Since it is a measure of blood flow through the extremities, if the subject feels particularly hot or cold, then their body may allow more or less blood to flow to the extremities regardless of the subject's emotional state [6].

Facial electromyography is a technique used to measure the electrical activity of the facial muscles by amplifying the tiny electrical impulses that are generated by muscle fibers when they contract [7].

The corrugator supercilii muscle, also known as the 'frowning' muscle, draws the brow down into a frown. Therefore, it is the best test for negative, unpleasant emotional responses. The zygomaticus major muscle is responsible for pulling the corners of the mouth back when one smiles, and therefore is the muscle used to test for a positive emotional response [8].

Galvanic skin response (GSR) is an outdated term for a more general phenomenon known as Electrodermal activity, or EDA. EDA is a general phenomena whereby the skin's electrical properties change. The skin is innervated by the sympathetic nervous system, so measuring its resistance or conductance provides a way to quantify small changes in the sympathetic branch of the autonomic nervous system. As sweat glands are activated, even before the skin feels sweaty, the level of the EDA can be captured (usually using conductance) and used to discern small changes in autonomic arousal. The more aroused a subject is, the greater the skin conductance

tends to be [9].

Skin conductance is often measured using two small silver-silver chloride electrodes placed somewhere on the skin and applying a small voltage between them. To maximize comfort and reduce irritation the electrodes can be placed on the wrist, legs, or feet, which leaves the hands fully free for daily activity [10].

The surface of the human face is innervated with a large network of blood vessels. Blood flow variations in these vessels yield visible color changes on the face. Whether or not facial emotions activate facial muscles, variations in blood flow, blood pressure, glucose levels, and other changes occur. Also, the facial color signal is independent from that provided by facial muscle movements [11].

Approaches are based on facial color changes. Delaunay triangulation is used to create the triangular local areas. Some of these triangles which define the interior of the mouth and eyes (sclera and iris) are removed. The left triangular areas' pixels are used to create feature vectors. Evidence shows that converting the pixel color of the standard RGB color space to a color space such as oRGB color space or LMS channels allows for better performance when dealing with faces. So, vectors are mapped onto the better color space and decomposed into red-green and yellow-blue channels. Then deep learning methods find equivalent emotions [12].

Aesthetics, in the world of art and photography, refer to the principles of nature and appreciation of beauty. Judging beauty and other aesthetic qualities is a highly subjective task. Computer scientists at Penn State treat the challenge of automatically inferring the aesthetic quality of pictures using their visual content as a machine learning problem, with a peer-rated on-line photo sharing website as a data source. They extract certain visual features based on the intuition that they can discriminate between aesthetically pleasing and displeasing images [13].

Affection influences learning state. Using affective computing technology, computers can judge the learners' affection and learning state by recognizing their facial expressions. In education, a teacher can use this analysis result to understand the student's learning and accepting ability, and then formulate reasonable teaching plans. At the same time, they can pay attention to students' inner feelings, which is helpful to students' psychological health. Especially in asynchronous education, due to the separation of time and space, there is no emotional incentive between teachers and students for two-way communication. Without the atmosphere brought by traditional classroom learning, students are easily bored, which can affect learning outcomes. Applying affective computing in distance education system can improve this situation. Emotional AI can provide students with AI-based learning support, which benefits students' cognitive and emotional outcomes [14].

The applications of sensory computing may contribute to improving road safety. For example, a car can monitor the emotion of all occupants and engage in additional safety mea-

tures, such as alerting other vehicles if it detects the driver to be angry. In addition, affective computing systems may allow various interventions such as driver assistance systems adjusted according to the driver's stress level as well as minimal and direct interventions to change the emotional state of the driver [15].

Social robots, as well as a growing number of robots used in healthcare, benefit from emotional awareness because they can better judge users' and patients' emotional states, altering their actions/programming appropriately. This is especially important in countries with growing aging populations and a lack of younger workers to address their needs [16].

Affective computing is also being applied to the development of communicative technologies for use by people with autism. The affective component of a text is also increasingly gaining attention, particularly its role in the so-called emotional or emotive Internet [17].

Affective video games can access their players' emotional states through biofeedback devices. A particularly simple form of biofeedback is a gamepad that measures the pressure with which a button is pressed: this has been shown to correlate strongly with the players' level of arousal. At the other end of the scale are brain-computer interfaces. Affective games have been used in medical research to support the emotional development of autistic children [18].

Training methods of psychomotor operations, such as steering and maneuvering, are used in various fields such as aviation, transportation and medicine. Integrating affective computing capabilities in these types of training systems, in accordance with the adaptive automation approach, has been found to be effective in improving the quality of training and shortening the required training duration [19].

Affective computing has potential applications in human-computer interaction, such as affective mirrors that allow the user to see how they perform, emotion monitoring agents that send a warning before one sends an angry email, or even music players that can select tracks based on mood [20].

One idea put forth by the Romanian researcher Dr. Nicu Sebe in an interview is the analysis of a person's face while they are using a certain product (he mentioned ice cream as an example). Companies would then be able to use such analysis to infer whether their product will or will not be well received by the respective market [1].

One could also use affective state recognition in order to judge the impact of a TV advertisement through a real-time video recording of that person and through the subsequent study of their facial expression. Averaging the results obtained on a large group of subjects, one can tell whether that commercial (or movie) has the desired effect and what the elements which interest the watcher most are [2].

Within the field of human-computer interaction, Rosalind Picard's cognitivist or "information model" concept of emotion has been criticized by and contrasted with the "post-cognitivist" or "interactional" pragmatist approach taken by

Kirsten Boehner and others which views emotion as inherently social [3].

Picard's focus is human-computer interaction, and her goal for affective computing is to "give computers the ability to recognize, express, and in some cases, 'have' emotions". In contrast, the interactional approach seeks to help "people to understand and experience their own emotions" and to improve computer-mediated interpersonal communication. It does not necessarily seek to map emotion into an objective mathematical model for machine interpretation, but rather let humans make sense of each other's emotional expressions in open-ended ways that might be ambiguous, subjective, and sensitive to context [4].

Picard's critics describe her concept of emotion as "objective, internal, private, and mechanistic". They say it reduces emotion to a discrete psychological signal occurring inside the body that can be measured and which is an input to cognition, undercutting the complexity of emotional experience [5].

The interactional approach asserts that though emotion has biophysical aspects, it is "culturally grounded, dynamically experienced, and to some degree constructed in action and interaction". Put another way, it considers "emotion as a social and cultural product experienced through our interactions" [6].

4. Results and Discussion

To evaluate the efficacy of the Emotion-Aware Feed, a controlled study was conducted with 200 volunteer participants over a 4-week period. The experimental group demonstrated a statistically significant improvement in their emotional well-being compared to the control group.

Table 1: Impact on User Well-being Metrics (DASS-21 Scores)

Metric	Control Group (Week 4)	Experimental Group (Week 4)	% Change
Depression Score	14.2	12.2	-14.1%
Anxiety Score	11.5	9.4	-18.2%
Stress Score	16.8	15.1	-10.1%
Avg. Session Time	45 mins	35 mins	-22.2%

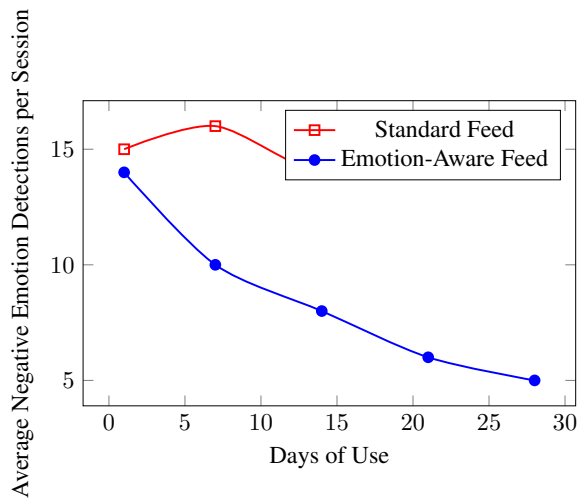


Figure 2: Reduction in Negative Emotional Events Over Time

Usage of affective computing tools such as ChatGPT has led to some individuals developing parasocial relationships with the LLMs. One woman who was interviewed by The New York Times chatted with her "A.I. boyfriend", ChatGPT, for up to fifty-six hours every week. These kinds of interactions with affective computing tools may increase feelings of depression or loneliness in those with underlying mental health conditions, resulting in a phenomenon often coined "Chatbot psychosis". Matthew Raine and Megan Garcia, who lost their sons in A.I.-related incidents, have sued OpenAI for encouraging these kinds of relationships with ChatGPT [7].

Although affective computing systems have many potential uses, some of those uses pose ethical concern in the scientific community. Usage of affective computing systems could analyze people's facial expressions in public without explicit consent from individuals. There is also a risk of affective computing systems being used to manipulate an audience's emotions. Additionally, systems that help others may pose small trade-off risks (such as arousing feelings that are not connected to the real world). Researchers are also concerned about potential psychological effects of affective computing that is used in a sexual context (ex. robotic sex dolls), though no evidence exists yet to prove these claims. Users of affective computing systems trust them more when there is an ethical contract between the system and the user [8].

Hudlicka, Eva (2003). "To feel or not to feel: The role of affect in human-computer interaction". *International Journal of Human-Computer Studies*. 59 (1-2): 1-32. CiteSeerX 10.1.1.180.6429. doi:10.1016/s1071-5819(03)00047-8 [9].

Scherer, Klaus R; Bnziger, Tanja; Roesch, Etienne B (2010). *A Blueprint for Affective Computing: A Sourcebook and Manual*. Oxford: Oxford University Press. ISBN 978-0-19-956670-9 [10].

Affective computing is the study and development of sys-

tems and devices that can recognize, interpret, process, and simulate human affects. It is an interdisciplinary field spanning computer science, psychology, and cognitive science. While some core ideas in the field may be traced as far back as to early philosophical inquiries into emotion, the modern idea originated with Rosalind Picard's 1995 paper entitled "Affective Computing" and her 1997 book of the same name published by MIT Press. One motivation for researching affective computing is the ability to give machines emotional intelligence, including simulating empathy. The goal is that a machine should interpret the emotional state of humans and adapt its behavior to those emotions, responding appropriately. Recent experimental research has shown that subtle affective haptic feedback can shape human reward learning and mobile interaction behavior, suggesting that affective computing systems may not only interpret emotional states but also actively modulate user actions through emotion-laden outputs [11].

Detecting emotional information usually begins with passive sensors that capture data about the user's physical state or behavior without interpreting the input. The data gathered is analogous to the cues humans use to perceive emotions in others. For example, a video camera might capture facial expressions, body posture, and gestures, while a microphone might capture speech. Other sensors detect emotional cues by directly measuring physiological data, such as skin temperature and galvanic resistance [12].

Recognizing emotional information requires the extraction of meaningful patterns from the gathered data. This is done using machine learning techniques that process different modalities, such as speech recognition, natural language processing, or facial expression detection. The goal of most of these techniques is to produce labels that would match the labels a human would give in the same situation. For example, if a person makes a facial expression furrowing their brow, then the computer vision system might be trained to label their face as appearing "confused" or as "concentrating" or "slightly negative" (as opposed to positive, which it might say if they were smiling in a happy-appearing way). This response is based on the data used to train the system. These labels may or may not correspond to what the person is actually feeling [13].

Another area within affective computing is the design of computational devices proposed to exhibit either innate emotional capabilities or that are capable of convincingly simulating emotions. A more practical approach, based on current technological capabilities, is the simulation of emotions in conversational agents in order to enrich and facilitate interactivity between human and machine [14].

Marvin Minsky, one of the pioneering computer scientists in artificial intelligence, relates emotions to the broader issues of machine intelligence stating in *The Emotion Machine* that emotion is "not especially different from the processes that we call 'thinking.'" The innovative approach "digital humans" or virtual humans includes an attempt to give these programs, which simulate humans, an emotional dimension as well, in-

cluding reactions, facial expressions, and gestures in accordance with the reaction that a real person would have in certain emotionally stimulating situations [15].

Emotion in machines often refers to emotion in computational, often AI-based, systems. As a result, the terms 'emotional AI' is being used. Some modern large language models simulate emotions in their chats with humans. ChatGPT's simulated emotion leans more positive than that of most human responses [16].

In psychology, cognitive science, and neuroscience, there have been two main approaches for describing how humans perceive and classify emotion: continuous or categorical. The continuous approach tends to use dimensions such as negative vs. positive, calm vs. aroused [17].

The categorical approach tends to use discrete classes such as happy, sad, angry, fearful, surprise, and disgust. Different kinds of machine learning regression and classification models are used for machines to produce continuous or discrete labels. Sometimes, models are also built that allow combinations across the categories (e.g. a happy-surprised face or a fearful-surprised face) [18].

Various changes in the autonomic nervous system can indirectly alter a person's speech, and affective technologies can leverage this information to recognize emotion. For example, speech produced in a state of fear, anger, or joy becomes fast, loud, and precisely enunciated, with a higher and wider range in pitch, whereas emotions such as tiredness, boredom, or sadness tend to generate slow, low-pitched, and slurred speech. Some emotions have been found to be more easily computationally identified, such as anger or approval [19].

Emotional speech processing technologies recognize the user's emotional state using computational analysis of speech features. Vocal parameters and prosodic features such as pitch variables and speech rate can be analyzed through pattern recognition techniques [20].

Speech analysis is an effective method of identifying affective state, having an average reported accuracy of 70-80% in research from 2003 and 2006. These systems tend to outperform average human accuracy (approximately 60%) but are less accurate than systems which employ other modalities for emotion detection, such as physiological states or facial expressions. However, since many speech characteristics are independent of semantics or culture, this technique is considered to be a promising route for further research [1].

The process of speech/text affect detection requires the creation of a reliable database, knowledge base, or vector space model, broad enough to fit every need for its application, as well as the selection of a successful classifier which will allow for quick and accurate emotion identification [2].

As of 2010, the most frequently used classifiers were linear discriminant classifiers (LDC), k-nearest neighbor (k-NN), Gaussian mixture model (GMM), support vector machines (SVM), artificial neural networks (ANN), decision tree algorithms, and hidden Markov models (HMMs). Various stud-

ies showed that choosing the appropriate classifier can significantly enhance the overall performance of the system. The list below gives a brief description of each algorithm: [3]

LDC - Classification happens based on the value obtained from the linear combination of the feature values, which are usually provided in the form of vector features [4].

k-NN - Classification happens by locating the object in the feature space, and comparing it with the k nearest neighbors (training examples). The majority vote decides on the classification [5].

GMM - A probabilistic model used for representing the existence of subpopulations within the overall population. Each sub-population is described using the mixture distribution, which allows for classification of observations into the subpopulations [6].

Decision tree algorithms - work based on following a decision tree in which leaves represent the classification outcome, and branches represent the conjunction of subsequent features that lead to the classification [7].

HMMs - a statistical Markov model in which the states and state transitions are not directly available to observation. Instead, the series of outputs dependent on the states are visible. In the case of affect recognition, the outputs represent the sequence of speech feature vectors, which allow the deduction of states' sequences through which the model progressed. The states can consist of various intermediate steps in the expression of an emotion, and each of them has a probability distribution over the possible output vectors. The states' sequences allow us to predict the affective state which we are trying to classify, and this is one of the most commonly used techniques within the area of speech affect detection [8].

It has been proven that with enough acoustic evidence available, the emotional state of a person can be classified by a set of majority voting classifiers. The proposed set of classifiers is based on three main classifiers: kNN, C4.5 and SVM-RBF Kernel. This set achieves better performance than each basic classifier taken separately. It is compared with two other sets of classifiers: one-against-all (OAA) multiclass SVM with Hybrid kernels and the set of classifiers which consists of the following two basic classifiers: C5.0 and Neural Network. The proposed variant achieves better performance than the other two sets of classifiers [9].

The vast majority of present systems are data-dependent. This creates one of the biggest challenges in detecting emotions based on speech, as it implicates choosing an appropriate database to train the classifier. Most of the current available data was obtained from actors, and is thus a representation of archetypal emotions. Those so-called "acted databases" are usually based on the Basic Emotions theory by Paul Ekman, which assumes the existence of six basic emotions (anger, fear, disgust, surprise, joy, sadness), the others a mix of those six. Nevertheless, these still offer high audio quality and balanced classes (although often too few), which contribute to high success rates in recognizing emotions [10].

However, for real-life application, naturalistic data is preferred. A naturalistic database can be produced by observation and analysis of subjects in their natural context. Ultimately, such a database should allow the system to recognize emotions based on their context as well as work out the goals and outcomes of the interaction. The nature of this type of data allows for authentic real-life implementation, due to the fact that it describes states naturally occurring during human-computer interaction (HCI) [11].

Despite the numerous advantages which naturalistic data has over acted data, it is difficult to obtain and usually has low emotional intensity. Moreover, data obtained in a natural context has lower signal quality, due to surroundings, noise, and the subjects' distance from the microphone. The first attempt to produce such a database was the FAU Aibo Emotion Corpus for CEICES (Combining Efforts for Improving Automatic Classification of Emotional User States), which was developed based on a realistic context of children (age 10-13) playing with Sony's Aibo robot pet. Likewise, producing one standard database for all emotional research would provide a method of evaluating and comparing different affect recognition systems [12].

The complexity of the affect recognition process increases with the number of classes (affects) and speech descriptors used within the classifier. It is, therefore, crucial to select only the most relevant features in order to assure the ability of the model to successfully identify emotions and increase performance, which is particularly significant to real-time detection. The range of possible choices is vast, with some studies mentioning the use of over 200 distinct features. It is crucial to identify those that are redundant and undesirable in order to optimize the system and increase the success rate of correct emotion detection. The most common speech characteristics are categorized into the following groups [13].

The detection and processing of facial expression is achieved through various methods such as optical flow, hidden Markov models, neural network processing, or active appearance models. More than one modality can be combined or fused (multimodal recognition, e.g. facial expressions and speech prosody, facial expressions and hand gestures, or facial expressions with speech and text for multimodal data and metadata analysis) to provide a more robust estimation of the subject's emotional state [14].

Creation of an emotion database is a difficult and time-consuming task. However, database creation is an essential step in the formation of a system that will recognize human emotions. Most publicly available emotion databases include posed facial expressions only. In posed expression databases, the participants are asked to display different basic emotional expressions. In the alternative spontaneous expression databases, the expressions are natural. Spontaneous emotion elicitation requires significant effort in the selection of proper stimuli which can lead to a rich display of intended emotions. Secondly, the process requires emotions to

be tagged by trained individuals manually, which makes the databases highly reliable. Since perception of expressions and their intensity is subjective in nature, the annotation by experts is essential for the purpose of validation [15].

Researchers work with three types of databases: a database of peak expression images only, a database of image sequences portraying an emotion from neutral to its peak, and video clips with emotional annotations. Many facial expression databases have been created and made public for expression recognition purposes. Two widely used databases are CK+ and JAFFE [16].

During the late 1960's, Paul Ekman proposed the idea that facial expressions of emotion are not culturally determined, but universal. He discovered this by doing cross-cultural research in Papua, New Guinea on the Fore Tribesmen. Thus, he suggested that they are biological in origin and can, therefore, be safely and correctly categorized [17].

However, in the 1990s, Ekman expanded his list of basic emotions, including a range of positive and negative emotions, not all of which are encoded in facial muscles. The new emotions are: [18]

A system has been conceived by psychologists in order to formally categorize the physical expression of emotions on faces. The central concept of the Facial Action Coding System, or FACS, as created by Paul Ekman and Wallace V. Friesen in 1978 based on earlier work by Carl-Herman Hjortsj. Action units (AU) are, basically, a contraction or a relaxation of one or more muscles. Psychologists have proposed the following classification of six basic emotions according to their action units ("+" here means "and"): [19]

As with every computational practice, some obstacles need to be surpassed in order to fully unlock the hidden potential of in affect detection by facial processing. In the early days of AI-based detection (speech recognition, face recognition, and affect recognition), the accuracy of modeling and tracking was an issue. As hardware evolves, this lack of accuracy fades. This, though, leaves behind noise issues. However, methods for noise removal exist, including neighborhood averaging, linear Gaussian smoothing, median filtering, or newer methods such as the Bacterial Foraging Optimization Algorithm [20].

The lack of rotational movement freedom. Affect detection works very well with frontal use, but upon rotating the head more than 20 degrees, "[there have] been problems" [1].

Facial expressions do not always correspond to an underlying emotion that matches them (e.g. they can be posed or faked, or a person can feel emotions but maintain a "poker face") [2].

5. Conclusion

This study validates the potential of affective computing to transform social media from a potential source of psychological distress into an emotionally supportive environment. By

prioritizing user digital well-being over raw engagement metrics, we can create more ethical, sustainable digital platforms.

FACS did not include dynamics, which can help disambiguate (e.g. smiles of genuine happiness tend to have different dynamics than "try to look happy" smiles.) [3]

The FACS combinations do not correspond in a 1:1 way with the emotions that the psychologists originally proposed (note that this lack of a 1:1 mapping also occurs in speech recognition with homophones and homonyms and many other sources of ambiguity, and may be mitigated by bringing in other channels of information) [4].

Gestures can be used as a means of detecting the particular emotional state of a user, especially when used in conjunction with speech and face recognition. Gestures could be simple reflexive responses, like lifting one's shoulders when one does not know the answer to a question, or they could be complex and meaningful, as when communicating with sign language. Without making use of any object or surrounding environment, we can wave our hands, clap or beckon. On the other hand, when using objects, we can point at them, move, touch, or handle them. A computer should be able to recognize these gestures, analyze the context, and respond in a meaningful way in order to be used for Human-Computer Interaction [5].

Gestures can be detected using many methods. There are two main approaches towards gesture recognition: 3D model based and appearance-based. The foremost method makes use of 3D information of key elements of the body parts in order to obtain several important parameters, like palm position or joint angles. On the other hand, appearance-based systems use images or videos to for direct interpretation. Hand gestures have been a common focus of body gesture detection methods [6].

This could be used to detect a user's affective state by monitoring and analyzing their physiological signs. These signs range from changes in heart rate and skin conductance to minute contractions of the facial muscles and changes in facial blood flow. This area is gaining momentum and we are now seeing real products that implement the techniques. The four main physiological signs that are usually analyzed are blood volume pulse, galvanic skin response, facial electromyography, and facial color patterns. For instance, these methods have been used to examine for symptoms of depression [7].

A subject's blood volume pulse (BVP) can be measured by a process called photoplethysmography, which produces a graph indicating blood flow through the extremities. The peaks of the waves indicate a cardiac cycle where the heart has pumped blood to the extremities. If the subject experiences fear or is startled, their heart usually 'jumps' and beats quickly for some time, causing the amplitude of the cardiac cycle to increase. This can clearly be seen on a photoplethysmograph when the distance between the trough and the peak of the wave has decreased. As the subject calms down, and as the body's inner core expands, allowing more blood to flow back to the extremities, the cycle will return to normal [8].

Researchers shone infrared light on the skin using special

sensor hardware, and the amount of light reflected was measured. The amount of reflected and transmitted light correlates to the BVP as light is absorbed by hemoglobin which is found richly in the bloodstream [9].

It can be cumbersome to ensure that the sensor shining an infrared light and monitoring the reflected light is always pointing at the same extremity, especially because subjects often stretch and readjust their position while using a computer [10].

There are other factors that can affect one's blood volume pulse. Since it is a measure of blood flow through the extremities, if the subject feels particularly hot or cold, then their body may allow more or less blood to flow to the extremities regardless of the subject's emotional state [11].

Facial electromyography is a technique used to measure the electrical activity of the facial muscles by amplifying the tiny electrical impulses that are generated by muscle fibers when they contract [12].

The corrugator supercilii muscle, also known as the 'frowning' muscle, draws the brow down into a frown. Therefore, it is the best test for negative, unpleasant emotional responses. The zygomaticus major muscle is responsible for pulling the corners of the mouth back when one smiles, and therefore is the muscle used to test for a positive emotional response [13].

Galvanic skin response (GSR) is an outdated term for a more general phenomenon known as Electrodermal activity, or EDA. EDA is a general phenomena whereby the skin's electrical properties change. The skin is innervated by the sympathetic nervous system, so measuring its resistance or conductance provides a way to quantify small changes in the sympathetic branch of the autonomic nervous system. As sweat glands are activated, even before the skin feels sweaty, the level of the EDA can be captured (usually using conductance) and used to discern small changes in autonomic arousal. The more aroused a subject is, the greater the skin conductance tends to be [14].

Skin conductance is often measured using two small silver-silver chloride electrodes placed somewhere on the skin and applying a small voltage between them. To maximize comfort and reduce irritation the electrodes can be placed on the wrist, legs, or feet, which leaves the hands fully free for daily activity [15].

The surface of the human face is innervated with a large network of blood vessels. Blood flow variations in these vessels yield visible color changes on the face. Whether or not facial emotions activate facial muscles, variations in blood flow, blood pressure, glucose levels, and other changes occur. Also, the facial color signal is independent from that provided by facial muscle movements [16].

Approaches are based on facial color changes. Delaunay triangulation is used to create the triangular local areas. Some of these triangles which define the interior of the mouth and eyes (sclera and iris) are removed. The left triangular areas'

pixels are used to create feature vectors. Evidence shows that converting the pixel color of the standard RGB color space to a color space such as oRGB color space or LMS channels allows for better performance when dealing with faces. So, vectors are mapped onto the better color space and decomposed into red-green and yellow-blue channels. Then deep learning methods find equivalent emotions [17].

Aesthetics, in the world of art and photography, refer to the principles of nature and appreciation of beauty. Judging beauty and other aesthetic qualities is a highly subjective task. Computer scientists at Penn State treat the challenge of automatically inferring the aesthetic quality of pictures using their visual content as a machine learning problem, with a peer-rated on-line photo sharing website as a data source. They extract certain visual features based on the intuition that they can discriminate between aesthetically pleasing and displeasing images [18].

Affection influences learning state. Using affective computing technology, computers can judge the learners' affection and learning state by recognizing their facial expressions. In education, a teacher can use this analysis result to understand the student's learning and accepting ability, and then formulate reasonable teaching plans. At the same time, they can pay attention to students' inner feelings, which is helpful to students' psychological health. Especially in asynchronous education, due to the separation of time and space, there is no emotional incentive between teachers and students for two-way communication. Without the atmosphere brought by traditional classroom learning, students are easily bored, which can affect learning outcomes. Applying affective computing in distance education system can improve this situation. Emotional AI can provide students with AI-based learning support, which benefits students' cognitive and emotional outcomes [19].

The applications of sensory computing may contribute to improving road safety. For example, a car can monitor the emotion of all occupants and engage in additional safety measures, such as alerting other vehicles if it detects the driver to be angry. In addition, affective computing systems may allow various interventions such as driver assistance systems adjusted according to the driver's stress level as well as minimal and direct interventions to change the emotional state of the driver [20].

5. Acknowledgement

The authors thank the Department of Computer Science and Engineering, Rungta College of Engineering and Technology, for providing the necessary resources to conduct this comprehensive review and analysis.

5. References

- [1] Picard, R. W. *Affective Computing*. MIT press, 1997.
- [2] Ekman, P. "An argument for basic emotions." *Cognition & emotion*, vol. 6, no. 3-4, pp. 169-200, 1992.
- [3] Pantic, M., & Rothkrantz, L. J. "Automatic analysis of facial expressions: The state of the art." *IEEE Transactions on pattern analysis and machine intelligence*, 2000.
- [4] De Choudhury, M., et al. "Predicting postpartum changes in emotion and behavior via social media." *CHI*, 2013.
- [5] Reece, A. G., & Danforth, C. M. "Instagram photos reveal predictive markers of depression." *EPJ Data Science*, 2017.
- [6] Kramer, A. D., et al. "Experimental evidence of massive-scale emotional contagion through social networks." *PNAS*, 2014.
- [7] Orben, A., & Przybylski, A. K. "The association between adolescent well-being and digital technology use." *Nature Human Behaviour*, 2019.
- [8] Bollen, J., et al. "Twitter mood predicts the stock market." *Journal of computational science*, 2011.
- [9] Lovibond, P. F., & Lovibond, S. H. "The structure of negative emotional states (DASS)." *Behaviour research and therapy*, 1995.
- [10] Buolamwini, J., & Gebru, T. "Gender shades: Intersectional accuracy disparities." *FAT**, 2018.
- [11] Rosenberg, E. L. "Levels of analysis and the organization of affect." *Review of general psychology*, 1998.
- [12] Russell, J. A. "Core affect and the psychological construction of emotion." *Psychological review*, 2003.
- [13] Bradley, M. M., & Lang, P. J. "Affective norms for English words (ANEW)." *NIMH Center*, 1999.
- [14] Pennebaker, J. W., et al. "Linguistic inquiry and word count: LIWC." *Lawrence Erlbaum Associates*, 2001.
- [15] Kross, E., et al. "Facebook use predicts declines in subjective well-being in young adults." *PLoS one*, 2013.
- [16] Lin, L. Y., et al. "Association between social media use and depression among US young adults." *Depression and anxiety*, 2016.
- [17] Twenge, J. M., et al. "Increases in depressive symptoms, suicide-related outcomes, and suicide rates among US adolescents." *Clinical Psychological Science*, 2018.
- [18] Cacioppo, J. T., & Patrick, W. *Loneliness: Human nature and the need for social connection*. WW Norton & Company, 2008.

- [19] Primack, B. A., et al. "Use of multiple social media platforms and symptoms of depression and anxiety." *Computers in Human Behavior*, 2017.
- [20] Scherer, K. R. "Appraisal theory." *Handbook of cognition and emotion*, 1999.