

Nurturing Emotions from Vision to Natural Language Processing

Y.Y. Nguwi

Nanyang Business School
Nanyang Technological University, Singapore
Singapore

Abstract— The crucial facets of human intelligence categorized by reasoning, problem solving and learning [1] have been well achieved by machine since the early days. The advancement in Large Language Model (LLM) demonstrated how machine achieved human's linguistic ability. Linguistic ability has driven active research in the area of natural language processing, partly motivated to pass The Turing Test [2]. Emotion recognition in machine, on the other hand, has been developed at a relatively slower pace. Are human emotions understood well in machine? In this study, we looked at emotion recognition from the perspective of computer vision and natural language. This paper presented the state of the art of emotion recognition achieved thus far, the mapping of deep learning models for emotion recognition and use cases of emotion recognition.

Keywords— emotion recognition, text, computer vision, ai.

I. INTRODUCTION

Computer was first invented to solve preprogramed calculation; the first general purpose computer can be traced back to the Analytical Engine designed by Charles Babbage [3] back in the year between 1833-1843. Ada Lovelace seeded the "intelligence" in machine with computer program and believed that computer would be able to do everything except thinking. Her translation notes were discovered by Alan Turing who later shaped this to form the launching point for artificial intelligence with the publication of "Computing Machinery and Intelligence" [4] and "Can Machine Think?" [2]. The first mention of "Artificial Intelligence" was made and defined by John McCarthy in the year 1956 [5].

Machine has achieved the ability to see, hear and speak. The ability to see is termed "Computer Vision", providing vision for computer to see and understand the objects in the world through pattern recognition. Computer vision builds on different image processing techniques to extract the object of interest for recognition. This has been an area with significant research and development in recent years as machine has not been built to understand patterns well. Rule-based approach does not work well as objects can appear differently in varying background and intermingled with other objects. Google's Large Language Model (LLM) has been trained with 100B parameters and the LLM like model for vision is currently at 22B parameters [6]. The ImageNet project [7] has collected more than 14 million of images and organized according to WordNet hierarchy. The use cases for computer vision are plentiful. In healthcare, some works achieved significant breakthrough ahead of other industries. For example, deep learning-based skin cancer detection [8-10] achieved dermatologist level of accuracy of as high as 98.3%, lung

cancer stage detection achieved pathologists' comparable accuracy of 97% [11], diagnosing COVID19 and other lung diseases detection [12], gastrointestinal cancer detection [13], breast cancer detection [14, 15], and IBM Watson prescribing cancer treatment for brain cancer patients [16, 17].

Another mature use case for computer vision is vision based smart vehicles to recognize objects on the road for smart steering. A smart vehicle combines several computer vision decision systems like road sign detection [18, 19], lane detection [20, 21], road scene segmentation [22], vehicle detection [23] and other smart system for assisted driving. Other potential extension includes automated car damage inspection system [24, 25] to automate car insurance claim [26].

Speech recognition starts with converting speech waveforms into spectrograms, followed by frequency transform and transcription through natural language processing. It has found use cases in virtual assistants, voice activated intelligent transport system, conversational AI to auto transcribe dialog, voice search and customer service auto responding system. The deep learning speech recognition system adopted Convolution Neural Network or Recurrent Neural Network [27]. Other Deep Learning system for Speech recognition includes Generative Adversarial Network (GAN) and Transfer Learning [28].

Emotion is a high-level task that represents human inner state. In order to understand emotions and react accordingly is vital in maintaining relationships. Affective computing relates to the study in computing field relates to human affection on the various emotional states. Human brain is primary originator of emotion which shows brainwave detected by Electroencephalogram (EEG) [29]. The EEG signal is recorded through different measure points over the scalp to understand the activation signal of neurons. EEG studies [30] have shown significant differences between positive and negative emotions. The left anterior region of the brain relates to positive emotion, and the right anterior region of brain relates to negative emotion. More accurate monitoring can be measured using Magneto Resonance Imaging (MRI) [31]. This complicated response has not been well modelled by machine. Further, emotion triggers response and connection between amygdala and hippocampal complex [32] which further complicates the possible modelling from existing brain inspired deep learning model. Darwin published the first facial expression work [33]. It highlighted that some emotions are inborn. A century later, Ekman and Friesen [34] published the foundational primary emotions which expressed six distinctive unique facial expressions.

Apart from understanding emotion through facial expression, tone of voice reflects human emotions as well. The acoustic characteristics of voice, such as pitch range, rhythm, amplitude, or duration variations, can offer indications of valence and specific emotions. [35]. A user feeling bored or sad would generally speak with a slower pace, lower pitch, and minimal high frequency energy in their speech. Conversely, someone experiencing fear, anger, or joy would speak at a faster rate, with louder volume and an emphasis on high frequency energy [36]. Recent deep learning work on voice emotion recognition utilized Convolution Neural Network [37] to model architecture of visual cortex with multiple cortex layers, amplitude and frequency of speech were used as features. The database of speech recognition generally consists of recorded speech synthesis, prosodic modelling, speech conversion that are voiced out by actors like the work in [38].

Understanding emotion from text is classified under text recognition. Calefeto et. al. published a toolkit for emotion recognition from nine thousand text banks [39] to recognize six basic emotions based on Support Vector Machines. Deep learning-based system can be found in [40] based on bidirectional Long Short Term Memory and Convolution Neural Network on the same standardized six basic emotions. Recognizing additional no standardized emotions like Anticipation and Trust were attempted by [41]. Overall, the recognition results vary depending on subprocesses within Natural Language Processing like Part of Speech tagging and parsing [42], the approaches for implicit and explicit emotion recognition were studied.

In this work, we studied the backgrounds of emotion recognition and the recent development. The current state of the art for emotion recognition and use cases presents opportunity for future works. This work ends with illustration the use cases of emotion recognition, how it can be used to advance the field of financial services, manufacturing, real estate, accommodation, and healthcare.

II. METHODOLOGY AND TOOLS

Emotion recognition in Artificial Intelligence was first started by Picard [43] and classified this under affective computing that attempted to bridge the understanding on human emotion from psychology to computing. It laid the groundwork for computers to recognize, interpret and express emotions. Several models were proposed to utilize images, speech, physiological signals and text data. Facial emotion and expressed text represent voluntary disclosure to the currently experienced emotion. Visual interpretation of emotion in computer vision goes through processes of locating face, image segmentation, region extraction and emotion classification. Text data expresses emotion directly or indirectly. Its interpretation goes through text encoding, machine translation, tokenization, root word extraction, tagging and classification. Emotion expressed through facial expression may not capture the entirety of human emotion as mixed emotion could be felt while only the more pronounced emotion is expressed. The response is also not universal, and is influenced by other factors like memory, past emotional experiences, and differences in individuals. Human emotion is processed in brain that coordinates facial muscles movement through complex network. Multiple parts of brains are linked to

emotion understanding and expression. For example, hippocampus encodes emotional memory and interact emotions [32, 44], amygdala process and generate emotion response [32, 44], prefrontal cortex regulates and modulates emotion response [45, 46], while cingulate cortex interacts and coordinates with other parts of brain and facial muscle to process emotion [47, 48]. Different parts of brains work in conjunction to regulate, comprehend, and express the emotion as response. When the brain receives emotional response like happy or sad, the relevant facial muscles are activated to express the corresponding facial expression. As such, architecture for neural networks should be of multiple facets for emotion understanding and response.

Linguistic ability is unique to human. Thoughts are processed in multiple parts of brains expressed in words that constitutes verbal expression. Language enables more varied ways to express emotions than facial expression as nonverbal expression. Amygdala plays a significant role in language and emotion, it integrates emotion into language comprehension and expression [49, 50]. Hippocampus takes care of processing emotion information and plays a key role in components of language [51] including expressing verbal emotion. Other related parts include Anterior Cingulate Cortex, Orbitofrontal Cortex, and Ventromedial Prefrontal Cortex work together and play critical role in emotion and decision making [52]. The role of language in emotion was found to be bidirectional where both affect and influence each other [53]. Language is also a crucial factor [53] in understanding emotion processes. Brooks et. al. [54] discussed how language influences the neural processes in the perception and experience of emotions. It suggested language acts as a mediator for emotion and shape the response with regulation. The labelling of emotion like happy or sad assists brain to recognize the emotion and hence reinforce the labelling process.

Deep learning network is the closest structure to model the abovementioned activities when emotion is encountered or expected. Neural network based model started from a single perceptron that can only solves linear problem, to feedforward neural network capable of solving nonlinear problem, it offers only single direction propagation of activation signal. Convolution Neural Network (CNN) works on the idea dissecting vision data into multiple segments or layers, it evolves from the idea of image convolution which multiplies segment of image for aggregation into another set of matrices. The layering approach was inspired by human visual cortex which maps to visuals using cascading segments of area similar to applying different filters on data. The origin of CNN architecture can be traced back to “neocognitron” by Fukushima [55] who implemented the idea of simple cell (Scell) and complex cell (Ccell) found in visual primary cortex in CNN. The architecture is further refined by including pooling layers to scale down the large input features and downsample to a subset that is representative of its original salient features. To overcome the problem of overfitting, the concept of Dropout [56] was introduced to randomly drop units during the training process to prevent neurons from coadaptation.

Machine learning process goes through a series of generalized steps like features representation, features engineering, followed by learning and performance measurements. The

desired emotion understanding framework consists of modules to understanding the currently experienced emotion followed by reaction to respond to the emotion. Recent works on emotion understanding can be categorized into the following main types:

A. Convolutional Neural Network

Convolutional Neural Network (CNN) is often used for image classification and video analysis. Hierarchy and relationship between features are learnt and convolved by the network. CNN can extract and distinguish features such as AlexNet [57], Residual Net (ResNet) [58, 59], and Visual Geometry Group (VGG) [6062]. The details of number of layers, parameters are provided in Table 1. The number of training images are similar at 1.2 million.

B. Recurrent Neural Network

Recurrent Neural Network (RNN) is designed to handle sequential data and using recurrent connection to retain data in the earlier time steps as hidden state. It is useful data expressed in time steps like speech, video, and text sequence data. It builds on the backpropagation algorithm to backpropagate through time to minimize the error in prediction. Long term dependency is handled well in variants like Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU). LSTM [63] started from addressing vanishing gradient issue in RNN due to using earlier data. [64] proposed advanced LSTM for weighted pooling RNN with accuracy of 55.3%. Semantic and emotional word vectors were adopted to enhance emotion recognition performance of LSTM in [65] with accuracy of 70.66%. The base GRU model was proposed in 2014 [66] introduced as an alternative to LSTM for machine translation task. The improvement is on lesser parameters required thus lower computations. Improved variant like Bidirectional GRU incorporates an attention mechanism was proposed [67] for text emotion recognition. The attention mechanism allows for selective focus on emotion related information. Emotion recognition from dialogue was studied in [68] with Convolutional Self-Attention Network (CANGRU) with baseline performance of 6567% and optimized to 8488% with pretrained model. Bidirectional models improved the models by approximately 2%.

Libraries for deep learning are actively being developed and updated. The earliest implementation, Torch, started from year 2002 was implemented in C Programming language. Followed by rapid development in TensorFlow, Keras and PyTorch since year 2015. Table 2 illustrates the main deep learning libraries with integration of Recurrent and Convolution Neural Networks. Table 3 tabulates workflow based deep learning tools where deep learning workflow can be designed with basic understanding of deep learning foundation. The tiers of licenses and deep learning libraries they integrate are shown. Table 4 lists down the libraries or SDK that comes with integration of emotion recognition. Some have been pretrained with billions of texts or image emotion data.

III. USE CASES FOR EMOTION UNDERSTANDING

In this section, we look at the use cases that can be developed from emotion recognition or understanding from visual or

text. The use cases are arranged by industries. There are some existing use cases and some untouched opportunities. According to the latest AI Index Report [69], research in AI continues to rise rapidly with industry races ahead of academia from year 2014 onwards due to larger amounts of data, resources and fund required in which industry actors inherently possess in greater amounts as compared to academia.

A. Financial Services

Fraud detection has been a long-standing priority in financial services. Fraud is costly for finance industry; it is estimated that every \$1 of fraud costs businesses \$4 [70]. The early detection of misconduct is enforced in some jurisdictions. Emotion recognition can provide additional cue to the likelihood of fraudulent transactions when client is performing the transaction. Emotion touch point can be over voice for telephone banking, or facial expressions recognition for transaction at physical branch. Questionnaires can be asked during online transactions to gather additional emotion cues. Endo et. al. [71] conducted experiments using voice recordings from individuals engaged in fraud activities, and it exhibited heightened emotional patterns like persuasion, excitement and anxiety. In another study, [72] uses text to predict financial distress using emotion understanding. There are cases of fraudulent transactions from internal employees, like the cases in JPMorgan Chase [73, 74] and other financial institution [75]. Emotion recognition at workplace can potentially provide early detection of unethical act before it is too late. Boyd et. al. [76] proposed ethical speculation lens to examine emotion at workplace.

B. Manufacturing

Combining other technology like Internet of Things (IoT) to embed emotion recognition can potentially transform manufacturing industries. Emotion recognition can be embedded in manufacturing processes for detecting sudden change in emotion responses of workers on the floor and pick up issues or defects. Monitoring of emotion responses can be a coherent source of input to improve overall productivity and improve product yield. Hu et. al. [77] proposed an iRobot-Factory that interconnects production line and incorporates cognitive emotion recognition of users to enhance decision making and problem solving in manufacturing systems. This idea can be further developed into manufacturing floor ergonomic redesign to layout machines, workstation for improved worker's comfort and fatigue reduction. It was found that fatigue is often associated with emotional exhaustion [78, 79], hence the use of emotion detection could provide an initial cues and further cues like alertness in eyes can provide useful features for fatigue detection. Future manufacturing can incorporate facial expressions in the process of manufacturing personalized products, for example the work in [80] proposed the idea of personalized jewelry manufactured from emotion of individuals.

C. Real Estate

Real Estate industry entails the development, leasing, marketing, management of residential, commercial, or industrial properties [81]. Most parts involve heavy human interactions, communication, and negotiation. The use of face recognition or emotion recognition system subjects to privacy

laws. The use of individual information, consent and anonymization should be addressed, this is discussed in [82, 83]. In a recent study [84] that examined the role of emotion as a response to property attributes, attributes like location and parking space display high significance from the inferred emotions. The relationship between emotion and renovation solution for a building was studied in [85], emotional state is affected and expressed in the building being renovated with sustainable design. Future work in this area can be extended to linking emotion to property viewing, from there understand factors that influence viewers and shortlist properties that fulfil the criteria. Emotion plays complementary role in helping buyers to filter properties of their likings.

D. Accommodation and Tourism

The use of emotion recognition in hotel industry was studied in [86] to enhance customer experience. Feedback system can incorporate emotion for issues to be attended to. Adaptive room environment to adjust room settings can create personalized experience based on the expressed emotions. Designing emotion-based tourism recommenders in wearable form was proposed in [87]. Multi-modal approach was used to construct customized user profiles and recommendations. Risk of destination is a challenge to be determined as future works. The incorporation of chatbot can also foster improved customer satisfaction [88] with emotion understanding. The other related works include service bots in hotel environments [89, 90], using sensors and provide service using the service bots. The emotions and expectations of guests are captured through the interaction with service bots. Barriers of adopting service bots include technological barriers especially for less tech savvy people, lack of human touch during the interaction, and complexity results in longer time spent to figure out how to use the bots [91].

E. Healthcare

Affective computing has great potential in healthcare industry. For example, in patient care, patient rehabilitation, and diagnosis. Understanding the emotions of patient in pain for patient care, patients recovering from surgery, and elderly who may not be able to express themselves well will improve the quality of care. Emotion recognition can provide more systematic way for various diagnosis to aid the decision of clinicians. Zhiyun et. al. [92] conducted a systematic review on the use of emotion recognition to identify schizophrenia, studies have shown impaired facial emotion are observed in patients with schizophrenia. Other major mental health also displays traces on distressed emotions like depressive disorder [93, 94]. Recent works adopting deep neural network for emotion recognition reported results in the range of 85% to detect cognitive impairment from facial emotion [95]. Using multiple sources of data like speech, text and EEG will further optimize the accuracy. Early work to attempt emotion recognition in military health care setting can be found in [96] to detect emotional change under stress for potential post-traumatic stress disorder, depression and suicide through voice. Kodati and Tene [97] attempted to detect negative emotions through patient text. Egger et. al. [98] reviewed different measures of ECG, EDA and other physiological signal to enhance emotion recognition from lab setting to wearable options.

IV. CONCLUSION

In this paper, we looked at the ways and means of systems that digitalizes and understands human emotions. The Artificial Intelligence progresses from the seeding question of "Can Machine Think?". The development in affective computing looks at how this can further progress to emotion exchange between human and machine. This work links the ability of machine to see, hear and speak to emotion understanding. It started off by looking at emotion as an outcome of mixed exchanges in different parts of brains to express it through facial expression and languages. This forms the early inspiration for convolution neural network to mimic the way vision and languages are processed by dissecting visuals or text being perceived followed by pooling and dropout to complete the recognition task. The state of the art of the relevant trained models, libraries, and workflow-based tools were also presented.

REFERENCES

- [1] Colom, R., et al., Human intelligence and brain networks. *Dialogues Clin Neurosci*, 2010. 12(4): p. 489-501.
- [2] Turing, A.M., Can a machine think. *Mind*, 1950. 59(236): p. 433-460.
- [3] Babbage, C., Babbage's analytical Engine. *Astronomische Nachrichten*, 1843. 21: p. 157.
- [4] Turing, A.M., Computing Machinery and Intelligence. *Creative Computing*, 1980. 6(1): p. 44-53.
- [5] McCarthy, J., et al., A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955. *AI magazine*, 2006. 27(4): p. 12-12.
- [6] Dehghani, M., et al., Scaling vision transformers to 22 billion parameters. *arXiv preprint arXiv:2302.05442*, 2023.
- [7] Deng, J., et al. Imagenet: A large-scale hierarchical image database. in 2009 IEEE conference on computer vision and pattern recognition. 2009. Ieee.
- [8] Esteva, A., et al., Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 2017. 542(7639): p. 115-118.
- [9] Sharma, A.K., et al., Dermatologist-level classification of skin cancer using cascaded ensembling of convolutional neural network and handcrafted features based deep neural network. *IEEE Access*, 2022. 10: p. 17920-17932.
- [10] Rezvantab, A., H. Safigholi, and S. Karimijeshni, Dermatologist level dermoscopy skin cancer classification using different deep learning convolutional neural networks algorithms. *arXiv preprint arXiv:1810.10348*, 2018.
- [11] Coudray, N., et al., Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nature Medicine*, 2018. 24(10): p. 1559-1567.
- [12] Ibrahim, D.M., N.M. Elshennawy, and A.M. Sarhan, Deep-chest: Multi-classification deep learning model for diagnosing COVID-19, pneumonia, and lung cancer chest diseases. *Computers in Biology and Medicine*, 2021. 132: p. 104348.
- [13] Kather, J.N., et al., Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nature medicine*, 2019. 25(7): p. 1054-1056.
- [14] Khan, S., et al., A novel deep learning based framework for the detection and classification of breast cancer using transfer learning. *Pattern Recognition Letters*, 2019. 125: p. 1-6.
- [15] Wang, D., et al., Deep learning for identifying metastatic breast cancer. *arXiv preprint arXiv:1606.05718*, 2016.
- [16] Ross, C. and I. Swetlitz, IBM's Watson supercomputer recommended 'unsafe and incorrect' cancer treatments, internal documents show. *Stat*, 2018. 25.
- [17] Strickland, E., IBM Watson makes a treatment plan for brain-cancer patient in 10 minutes; doctors take 160 hours. *IEEE Spectrum*, Posted, 2017. 11.

- [18] Tabernik, D. and D. Skočaj, Deep learning for large-scale traffic-sign detection and recognition. *IEEE transactions on intelligent transportation systems*, 2019. 21(4): p. 1427-1440.
- [19] Liu, C., et al., Machine vision based traffic sign detection methods: review, analyses and perspectives. *IEEE Access*, 2019. 7: p. 86578-86596.
- [20] Bar Hillel, A., et al., Recent progress in road and lane detection: a survey. *Machine vision and applications*, 2014. 25(3): p. 727-745.
- [21] Kim, Z., Robust lane detection and tracking in challenging scenarios. *IEEE Transactions on intelligent transportation systems*, 2008. 9(1): p. 16-26.
- [22] Kherraki, A., M. Maqbool, and R.E. Ouazzani. Lightweight and Efficient Convolutional Neural Network for Road Scene Semantic Segmentation. in 2022 IEEE 18th International Conference on Intelligent Computer Communication and Processing (ICCP). 2022.
- [23] Wu, R., et al. Real-time Vehicle Detection System for Intelligent Transportation using Machine Learning. in 2022 IEEE Green Energy and Smart System Systems (IGESSC). 2022.
- [24] Patil, K., et al. Deep learning based car damage classification. in 2017 16th IEEE international conference on machine learning and applications (ICMLA). 2017. IEEE.
- [25] Dhieb, N., et al. A very deep transfer learning model for vehicle damage detection and localization. in 2019 31st international conference on microelectronics (ICM). 2019. IEEE.
- [26] Singh, R., et al. Automating car insurance claims using deep learning techniques. in 2019 IEEE fifth international conference on multimedia big data (BigMM). 2019. IEEE.
- [27] Soni, S., R.N. Yadav, and L. Gupta, State-of-the-Art Analysis of Deep Learning-Based Monaural Speech Source Separation Techniques. *IEEE Access*, 2023. 11: p. 4242-4269.
- [28] Minaee, S., et al., Biometrics recognition using deep learning: a survey. *Artificial Intelligence Review*, 2023.
- [29] Sears, A. and J.A. Jacko, *Human-computer interaction fundamentals*. 2009: CRC press.
- [30] Harmer, M., R. EW, and B. RJ, Rapid sintering of pure and doped alpha-Al₂O₃. 1979.
- [31] Westen, D., et al., Neural bases of motivated reasoning: An fMRI study of emotional constraints on partisan political judgment in the 2004 US presidential election. *Journal of cognitive neuroscience*, 2006. 18(11): p. 1947-1958.
- [32] Phelps, E.A., Human emotion and memory: interactions of the amygdala and hippocampal complex. *Current opinion in neurobiology*, 2004. 14(2): p. 198-202.
- [33] Hess, U. and P. Thibault, Darwin and emotion expression. *American psychologist*, 2009. 64(2): p. 120.
- [34] Ekman, P. and W.V. Friesen, Constants across cultures in the face and emotion. *Journal of personality and social psychology*, 1971. 17(2): p. 124.
- [35] Ball, G. and J. Breese, Emotion and personality in a conversational agent. *Embodied conversational agents*, 2000. 189.
- [36] Picard, R.W. and J. Healey, Affective wearables. *Personal technologies*, 1997. 1: p. 231-240.
- [37] Alu, D., E. Zoltan, and I.C. Stoica, Voice based emotion recognition with convolutional neural networks for companion robots. *Science and Technology*, 2017. 20(3): p. 222-240.
- [38] Barra Chicote, R., et al., Spanish expressive voices: Corpus for emotion research in spanish. 2008.
- [39] Calefato, F., F. Lanubile, and N. Novielli. Emotxt: a toolkit for emotion recognition from text. in 2017 seventh international conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW). 2017. IEEE.
- [40] Batbaatar, E., M. Li, and K.H. Ryu, Semantic-emotion neural network for emotion recognition from text. *IEEE access*, 2019. 7: p. 111866-111878.
- [41] Park, S.-H., B.-C. Bae, and Y.-G. Cheong. Emotion recognition from text stories using an emotion embedding model. in 2020 IEEE international conference on big data and smart computing (BigComp). 2020. IEEE.
- [42] Alswaidan, N. and M.E.B. Menai, A survey of state-of-the-art approaches for emotion recognition in text. *Knowledge and Information Systems*, 2020. 62: p. 2937-2987.
- [43] Picard, R.W., *Affective computing*. 2000: MIT press.
- [44] Richardson, M.P., B.A. Strange, and R.J. Dolan, Encoding of emotional memories depends on amygdala and hippocampus and their interactions. *Nature neuroscience*, 2004. 7(3): p. 278-285.
- [45] Dixon, M.L., et al., Emotion and the prefrontal cortex: An integrative review. *Psychological bulletin*, 2017. 143(10): p. 1033.
- [46] Salzman, C.D. and S. Fusi, Emotion, cognition, and mental state representation in amygdala and prefrontal cortex. *Annual review of neuroscience*, 2010. 33: p. 173-202.
- [47] Stevens, F.L., R.A. Hurley, and K.H. Taber, Anterior cingulate cortex: unique role in cognition and emotion. *The Journal of neuropsychiatry and clinical neurosciences*, 2011. 23(2): p. 121-125.
- [48] Etkin, A., et al., Resolving emotional conflict: a role for the rostral anterior cingulate cortex in modulating activity in the amygdala. *Neuron*, 2006. 51(6): p. 871-882.
- [49] Ferstl, E.C., M. Rinck, and D.Y.v. Cramon, Emotional and temporal aspects of situation model processing during text comprehension: An event-related fMRI study. *Journal of cognitive Neuroscience*, 2005. 17(5): p. 724-739.
- [50] Lindquist, K.A., *Language and Emotion: Introduction to the Special Issue*. *Affective Science*, 2021. 2(2): p. 91-98.
- [51] Duff, M.C. and S. Brown-Schmidt, The hippocampus and the flexible use and processing of language. *Frontiers in human neuroscience*, 2012. 6: p. 69.
- [52] Rolls, E.T., et al., The human orbitofrontal cortex, vmPFC, and anterior cingulate cortex effective connectome: emotion, memory, and action. *Cerebral cortex*, 2023. 33(2): p. 330-356.
- [53] Lindquist, K.A., The role of language in emotion: existing evidence and future directions. *Current opinion in psychology*, 2017. 17: p. 135-139.
- [54] Brooks, J.A., et al., The role of language in the experience and perception of emotion: A neuroimaging meta-analysis. *Social Cognitive and Affective Neuroscience*, 2017. 12(2): p. 169-183.
- [55] Fukushima, K., Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological cybernetics*, 1980. 36(4): p. 193-202.
- [56] Srivastava, N., et al., Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 2014. 15(1): p. 1929-1958.
- [57] Krizhevsky, A., I. Sutskever, and G.E. Hinton, Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 2017. 60(6): p. 84-90.
- [58] Li, B. and D. Lima, Facial expression recognition via ResNet-50. *International Journal of Cognitive Computing in Engineering*, 2021. 2: p. 57-64.
- [59] He, K., et al. Deep residual learning for image recognition. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
- [60] Fathallah, A., L. Abdi, and A. Douik. Facial expression recognition via deep learning. in 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA). 2017. IEEE.
- [61] Cheng, S. and G. Zhou, Facial expression recognition method based on improved VGG convolutional neural network. *International Journal of Pattern Recognition and Artificial Intelligence*, 2020. 34(07): p. 2056003.
- [62] Simonyan, K. and A. Zisserman, Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [63] Hochreiter, S. and J. Schmidhuber, Long short-term memory. *Neural computation*, 1997. 9(8): p. 1735-1780.
- [64] Tao, F. and G. Liu. Advanced LSTM: A Study About Better Time Dependency Modeling in Emotion Recognition. in 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2018.
- [65] Su, M.-H., et al. LSTM-based text emotion recognition using semantic and emotional word vectors. in 2018 First Asian Conference on Affective Computing and Intelligent Interaction (ACII Asia). 2018. IEEE.
- [66] Cho, K., et al., Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.
- [67] Liu, T., Y. Du, and Q. Zhou, Text Emotion Recognition Using GRU Neural Network with Attention Mechanism and Emoticon Emotions, in *Proceedings of the 2020 2nd International Conference on Robotics, Intelligent Control and Artificial Intelligence*. 2020, Association for Computing Machinery: Shanghai, China. p. 278-282.
- [68] Jiang, T., et al. CAN-GRU: A Hierarchical Model for Emotion Recognition in Dialogue. 2020. Cham: Springer International Publishing.
- [69] Nestor Maslej, L.F., Erik Brynjolfsson, John Etchemendy, Katrina Ligett, Terah Lyons., H.N. James Manyika, Juan Carlos Niebles, Vanessa Parli, Yoav Shoham, Russell Wald, Jack Clark., and a.R. Perrault, *AI Index Report 2023 – Artificial Intelligence Index*. 2023.

- [70] Solution, L.R., Fraud Costs Increased More Than 10% over Pre-Pandemic Levels for APAC Businesses, According to LexisNexis Risk Solutions Study, in PR Newswire. 2022.
- [71] Endo, S. and K. Tanida, A Study on Voice Analysis Using Emotion Recognition for Detection of Special Fraud. IEICE Technical Report; IEICE Tech. Rep., 2020. 120(257): p. 14-19.
- [72] Hajek, P. and M. Munk, Speech emotion recognition and text sentiment analysis for financial distress prediction. Neural Computing and Applications, 2023: p. 1-15.
- [73] Katz, B., Three ex-Chase bank workers admit \$4.8 mln tax fraud, in Reuters. 2012.
- [74] Calle, J.P., JP Morgan Chase's financial troubles, in Juan Pablo Calle. 2022.
- [75] Biase, N., Former Bank Employee Charged With Million-Dollar Fraud And Embezzlement Scheme, in US Attorney's Office. 2023.
- [76] Boyd, K.L. and N. Andalibi, Automated emotion recognition in the workplace: How proposed technologies reveal potential futures of work. Proceedings of the ACM on Human-Computer Interaction, 2023. 7(CSCW1): p. 1-37.
- [77] Hu, L., et al., iRobot-Factory: An intelligent robot factory based on cognitive manufacturing and edge computing. Future Generation Computer Systems, 2019. 90: p. 569-577.
- [78] Liu, Y., et al. EEG-based evaluation of mental fatigue using machine learning algorithms. in 2018 International Conference on Cyberworlds (CW). 2018. IEEE.
- [79] Lewis, G. and S. Wessely, The epidemiology of fatigue: more questions than answers. Journal of epidemiology and community health, 1992. 46(2): p. 92.
- [80] Bertacchini, F., et al., Modeling and recognition of emotions in manufacturing. International Journal on Interactive Design and Manufacturing (IJIDeM), 2022. 16(4): p. 1357-1370.
- [81] Jackson, C., All About the Real Estate Industry, in C21. 2020.
- [82] Lewinski, P., J. Trzaskowski, and J. Luzak, Face and emotion recognition on commercial property under EU data protection law. Psychology & Marketing, 2016. 33(9): p. 729-746.
- [83] Naker, S. and D. Greenbaum, Now you see me: Now you still do: Facial recognition technology and the growing lack of privacy. BUJ Sci. & Tech. L., 2017. 23: p. 88.
- [84] Renigier-Bilozor, M., et al., Human emotion recognition in the significance assessment of property attributes. Journal of Housing and the Built Environment, 2022. 37(1): p. 23-56.
- [85] Velykorusova, A., et al., Intelligent Multi-Criteria Decision Support for Renovation Solutions for a Building Based on Emotion Recognition by Applying the COPRAS Method and BIM Integration. Applied Sciences, 2023. 13(9): p. 5453.
- [86] Marín-Morales, J., et al., Affective computing in virtual reality: emotion recognition from brain and heartbeat dynamics using wearable sensors. Scientific reports, 2018. 8(1): p. 13657.
- [87] Santamaria-Granados, L., J.F. Mendoza-Moreno, and G. Ramirez-Gonzalez, Tourist Recommender Systems Based on Emotion Recognition—A Scientometric Review. Future Internet, 2021. 13(1): p. 2.
- [88] Capsmart. Smart technologies in hospitality: A case study about chatbot usage. 2021.
- [89] Zalama, E., et al. Sacarino, a service robot in a hotel environment. in ROBOT2013: First Iberian Robotics Conference: Advances in Robotics, Vol. 2. 2014. Springer.
- [90] Pinillos, R., et al., Long-term assessment of a service robot in a hotel environment. Robotics and Autonomous Systems, 2016. 79: p. 40-57.
- [91] Wang, X., et al., Consumer resistance to service robots at the hotel front desk: A mixed-methods research. Tourism Management Perspectives, 2023. 46: p. 101074.
- [92] Gao, Z., et al., Facial Emotion Recognition in Schizophrenia. Frontiers in Psychiatry, 2021. 12.
- [93] Dalili, M.N., et al., Meta-analysis of emotion recognition deficits in major depressive disorder. Psychological medicine, 2015. 45(6): p. 1135-1144.
- [94] Dementescu, L.R., et al., Impaired attribution of emotion to facial expressions in anxiety and major depression. PloS one, 2010. 5(12): p. e15058.
- [95] Fei, Z., et al., Deep convolution network based emotion analysis towards mental health care. Neurocomputing, 2020. 388: p. 212-227.
- [96] Tokuno, S., et al. Usage of emotion recognition in military health care. in 2011 Defense Science Research Conference and Expo (DSR). 2011.
- [97] Dheeraj, K. and T. Ramakrishnudu, Negative emotions detection on online mental-health related patients texts using the deep learning with MHA-BCNN model. Expert Systems with Applications, 2021. 182: p. 115265.
- [98] Egger, M., M. Ley, and S. Hanke, Emotion Recognition from Physiological Signal Analysis: A Review. Electronic Notes in Theoretical Computer Science, 2019. 343: p. 35-55..

Table 1 Comparison of Convolutional Neural Network

	Variants	No. of Layers	No. of Parameters
AlexNet (Krizhevsky et al., 2017)	Conv1	96	349,440
	Conv2	256	614,656
	Conv3	384	885,120
	Conv4	384	1,327,872
	Conv5	256	884,992
Residual Net (He et al., 2016)	ResNet18	18	11 million
	ResNet34	34	21 million
	ResNet50	50	25.6 million
	ResNet101	101	44.5 million
	ResNet152	152	60.2 million
VGG (Simonyan & Zisserman, 2014)	VGG11	11	132 million
	VGG13	13	133 million
	VGG16	16	138 million
	VGG19	19	143 million

Table 2 Deep Learning Libraries for RNN and CNN

Library (Release Year)	RNN	CNN	Dependencies	Available from:
TensorFlow (2015)	✓	✓		https://www.tensorflow.org/
Keras (2015)	✓	✓	TensorFlow, Theano	https://keras.io/
PyTorch (2016)	✓	✓	Torch	https://pytorch.org/
Caffe (2013)		✓		https://caffe.berkeleyvision.org/
Theano (2007)	✓	✓		https://pypi.org/project/Theano/
Torch (2002)	✓	✓		https://pypi.org/project/torch/
MXNet (2015)	✓	✓		https://mxnet.apache.org/
FastAI (2018)	✓	✓	PyTorch	https://www.fast.ai/
TFLearn (2016)	✓	✓	TensorFlow	http://tflearn.org/
Lasagne (2014)	✓	✓	Theano	https://github.com/Lasagne/Lasagne
Deeplearning4j (2014)	✓	✓		https://deeplearning4j.konduit.ai/

Table 3 Workflow-based Deep Learning Tools

Tool	RNN	CNN	License
KNIME	✓	✓	Free + Paid Versions
WEKA	✓	✓	Free
RapidMiner	✓	✓	Free + Paid Versions
Orange	✓	✓	Free
Dataiku DSS	✓	✓	Free + Paid Versions
Alteryx	✓	✓	Free Trial + Paid Versions
Azure ML	✓	✓	Free + Paid Versions
H2O.ai	✓	✓	Free
IBM Watson Studio	✓	✓	Free + Paid Versions
Google AutoML	✓	✓	Paid Version
Databricks	✓	✓	Free + Paid Versions

Table 4 Libraries/SDK with Integration of Emotion Recognition

Library/SDK	RNN	CNN	Commercial
PyTorch Geometric		✓	No
TensorFlow.js	✓	✓	No
Affectiva SDK	✓	✓	Yes
Microsoft CNTK	✓	✓	Yes
OpenFace	✓	✓	No
EmoVoice			No
EmoPv	✓		No
DeepMojii	✓	✓	No