Applied Information Technology And Computer Science Vol. 4 No. 1 (2023) 1625-1636

AITCS

Homepage: http://publisher.uthm.edu.my/periodicals/index.php/aitcs
e-ISSN :2773-5141

Advanced Convolutional Neural Network for Accurate Detection of Different Facial Expression

Noor Aliah Natassyah Rhamie[1], Muhamed Saeed Jawad[1]*

[1]Faculty of Computer Science and Information Technology,
Universiti Tun Hussein Onn Malaysia, Parit Raja, Batu Pahat, 86400, MALAYSIA

*Corresponding Author Designation

DOI: https://doi.org/10.30880/aitcs.2023.04.01.098
Received 16 June 2022; Accepted 12 June 2023; Available online 30 June 2023

Abstract: Facial emotions is an expression that people always express in their daily communication such as happy face, sad face, stress face, angry face and confuse face. However, in this covid-19, people cannot see each other and they are separated from family and friends. Thus, this situation can lead into depression and anxiety in some people. Therefore, there is a development of artificial intelligent solution for detection and classification of facial emotions. People can recognize others emotions and can detect whether they in problems or need help. Pipeline methodology is used as the methodology to develop this application. This application was developed in window platform, from the testing had been done it had success to detect and classify human facial expression into emotions. This application is expected to ease educators to identify students' emotions so that they can undergo teaching and learning process successfully. It will also ease people to recognize other people's facial expressions and emotions

Keywords: Facial detection, Pipeline, Emotion

1. Introduction

Many companies these days want employees with a variety of courses and abilities. Typically, a worker will go through an interviewing procedure before being hired to work for a company. In an organization, time management and emotion are crucial components that will be examined. However, determining other people's emotions and moods without seeing their faces or expressions is difficult.

According to statistics from Berita Harian, the number of persons who commit suicide in Malaysia is on the rise. "There was 638 commit suicide cases that has been reported from January until July for this year" [1]. Director-General of health, Tan Sri Dr Noor Hisham Abdullah said that the number increased 1.4 times than last year cases which is 262 cases only. This is because to Covid-19's position, which prevents individuals from meeting close friends and most people are experiencing financial difficulties. Due to stress and anxiety issues, they prefer to remain silent and avoid contact with others.

*Corresponding author: saeed@uthm.edu.my

With this approach, which has been made easier by technological advancements in recent years, it can assist a variety of industries in improving education or manufacturing development. People, regardless of gender or age, will experience tension and worry. This can result in the worst-case scenario, such as suicide or self-harm.

People were previously unaware that they were suffering from sadness or stress overload. The majority of them are unconcerned about it and do not damage anyone. This is due to the fact that most of us are ignorant about depression.

Development of Artificial Intelligent Solution for Detection and Classification of Facial Emotions is one of the best steps that promises the best for the user. It can help people identify their own emotions and those around them more easily. It is also supposed to make it easier for educators to identify students' emotions so that they can go through the teaching and learning process successfully. Employers will gain from this project since it will ensure that workers' emotions are steady, resulting in a controlled working environment and a positive company reputation. This technology will make it easier for people to recognize other people's facial expressions and emotions [2].

The result of this facial emotion classification is to assist educators in identifying pupils who have emotional issues such as stress and despair. Employers can also rest assured that their employees will work cheerfully and provide a pleasant working atmosphere. By monitoring people's emotions, this device can indirectly identify persons who have been stressed. Aside from that, this technique can assist people who are experiencing uncontrollable emotions. Because most people die by suicide as a result of stress overload, it is critical that they take care of their own mental health.

There are five sections in this article. The first part is and introduction that describe the research context, namely facial emotion detection and classification with artificial intelligent solution. Related work is described in the second section. Research methodology and solution methods are discussed in the third section. The fourth section discusses the results and its analysis. Conclusion are presented in the final section

2. Related Work

2.1 Facial Emotion Detection

The artificial neuron is act like a human neuron [3]. Human brain learns by creating connections among these neurons. The multi-layer perception network is to connect multiple of these neurons in a multi-layer fashion. The more hidden layers, the more” deep” the network will get. Humans have specific social and emotional abilities that enable them engage with other people; one of these abilities is the ability to recognize the emotions of others. This skill significantly improves human connections.

As we move closer to a future with more Human-Machine Interaction systems, we must work to teach our robots to accomplish so-called social and emotional capacities using algorithms [4]. Recognizing emotion via face photos is a key research field in emotion detection. Recognizing a human's emotions allows the machine to change and tune itself to the human's requirements and comfort. Emotion detection can be done utilizing a variety of modalities, including video, audio, pictures, text, biometric data, and so on.

This research looks on the latest developments in emotion recognition using facial photos. Automated recognition of people's emotions has the ability to detect latent mental health disorders or predict latent mental health concerns in humans.

Over the last few decades, real-time emotion identification has been a hot topic of study. This project aims to classify the emotional expressions of a person that are disable such as deaf, bedridden, dumb and Autism children. Convolutional neural networks (CNN) and long short-term memory (LSTM) classifiers are used to detect emotions based on face key points and electroencephalograph (EEG) inputs [5]. Virtual markers were used to create an algorithm for real-time emotion recognition utilizing an optical flow method that works well under uneven lighting and subject head rotation up to 2 degrees.

Ten virtual markers are used to capture six facial emotions (joy, sadness, anger, fear, disgust, and surprise). Fifty-five undergraduate students (35 males and 25 females) with an average age of 22.9 years volunteered for the facial expression recognition experiment. EEG signals were collected by 19 undergraduate students who volunteered. Initially, Haar-like traits are employed to detect faces and eyes. After that, using the mathematical model technique, virtual markers are set on specific key points on the subject's face. The Lucas-Kande optical flow algorithm is used to track the markers.

For facial expression categorization, the distance between the subject's center of face and each marker position is employed as a feature. This distance attribute is statistically validated using a one-way analysis of variance with a significance level of p 0.01. The fourteen signals obtained from the EEG signal reader (EPOC+) channels are also used as EEG signal emotional categorization characteristics. Before being passed to the LSTM and CNN classifiers, the features are fivefold cross-validated.

Utilizing CNN for emotion detection using facial landmarks, it attained a rate of recognition emotion as maximum as 99.81 percent. However, for emotion identification using EEG signals, recognition rate that maximum it can attained using the LSTM classifier is 87.25 percent.

2.2 Emotion detection in artificial intelligent

Over the last few decades, real-time emotion identification has been a hot topic of study. This project aims to classify the emotional expressions of a person that are disable such as deaf, bedridden, dumb and Autism children. Convolutional neural networks (CNN) and long short-term memory (LSTM) classifiers are used to detect emotions based on face key points and electroencephalograph (EEG) inputs. Virtual markers were used to create an algorithm for real-time emotion recognition utilizing an optical flow method that works well under uneven lighting and subject head rotation up to 2 degrees [6].

Ten virtual markers are used to capture six facial emotions (joy, sadness, anger, fear, disgust, and surprise). Fifty-five undergraduate students (35 males and 25 females) with an average age of 22.9 years volunteered for the facial expression recognition experiment [7]. EEG signals were collected by 19 undergraduate students who volunteered. Initially, Haar-like traits are employed to detect faces and eyes. After that, using the mathematical model technique, virtual markers are set on specific key points on the subject's face. The Lucas-Kande optical flow algorithm is used to track the markers.

For facial expression categorization, the distance between the subject's center of face and each marker position is employed as a feature. This distance attribute is statistically validated using a one-way analysis of variance with a significance level of p 0.01. The fourteen signals obtained from the EEG signal reader (EPOC+) channels are also used as EEG signal emotional categorization characteristics. Before being passed to the LSTM and CNN classifiers, the features are fivefold cross-validated.

Utilizing CNN for emotion detection using facial landmarks, it attained a rate of recognition emotion as maximum as 99.81 percent. However, for emotion identification using EEG signals, recognition rate that maximum it can attained using the LSTM classifier is 87.25 percent.

2.2.1 Healthcare

By monitoring facial expressions in the waiting area, healthcare practitioners can utilize emotion recognition AI to prioritize their patients, especially in urgent care center where people don't make appointments. Those who are in most pain may be given top priority, while those with minor diseases may have to wait for an opening.

Researchers, on the other hand, are experimenting with emotion recognition technology. Researchers at Stanford University School of Medicine used Google Glass and a special smartphone app to assist autistic children better identify the emotions and facial expressions they saw. As long as the child is wearing Google Glass, the software provides real-time feedback on other people's facial expressions [8].

2.2.2 Marketing

Before releasing a product, marketers usually want to know how it will perform, but this is easier said than done. Focus groups aren't always as dependable as firms think, and some people are overly kind, afraid of hurting someone's feelings, so they sugarcoat their opinions rather than being candid. However, by evaluating the facial expressions of testers using their products or viewing their advertising, emotion recognition technology can help organizations get more out of their focus groups [8].

2.2.3 Manufacturer

Automobile makers can benefit from emotion recognition technologies as well. Cars that warn drivers when they are nodding off or becoming tired can help avoid dangerous collisions. The alarm could potentially be triggered by road rage or other intense emotions. This could be especially useful in cars that have self-driving or autopilot capabilities. The autopilot can engage if the human driver becomes extremely emotional or tired, while notifying the driver. The alarm may shock them awake or let them relax for a few moments before regaining control of the car [8].

2.3 Comparative study

2.3.1 Comparison of each technologies

A comparison of the previously discussed technologies is shown in Table 1. Due to the unavailability of trial versions or demos, nViso has not been tested. In terms of the outcomes, every technology that was examined demonstrated a high level of accuracy. However, various factors (reflection on glasses, poor lighting) obscure essential facial landmarks, resulting in inaccurate results. These technologies can identify an expression of suffering in a circumstance when the eyes and brows is unseen. As the example, when the subject is smile, it can be the stretching of mouth when it was opened.

In terms of speed, Emotion API and Affectiva both take about the same amount of time to scan an image. However, Kairos takes substantially longer. Furthermore, the number of values supplied by Affectiva gives developers a lot more information, making it easier to identify the emotion that the user is displaying than if we just have the weight of six emotions, for example [9]. Due to the availability of Affectiva, it provides free services to people dedicated to research and education or who produce less than $1,000,000 annually, is especially noteworthy.

Table 1 Comparison of emotion detection technologies from facial expression

Name	API/ SDK	Requires Internet	Information returned	Difficulty of use	Free Software
Emotion API	API/ SDK	Yes	Happiness, sadness, fear, anger, surprise, neutral, disgust, contempt	Low	Yes (Limited)

Affectiva	API/ SDK	Yes	Joy, sadness, disgust, contempt, anger, fear, surprise	Low	Yes, with some restriction
nViso	API/ SDK	No	Happiness, sadness, fear, anger, surprise, disgust and neutral	-	No
Kairos	API/ SDK	Yes	Anger, disgust, fear, joy, sadness, surprise	Low	Yes, only for personal use
Votakuri	SDK	No	Happiness, neutrality, sadness, anger and fear	Medium	Yes
Beyond Verbal	API	Yes	Temper Arousal Valence Mood (up to 432 emotions)	Low	No

3. Methodology/Framework

Pipeline methodology was used in this study. There are four steps in this procedure. Pre-processing, learning, evaluation and prediction are the four primary phases of the Machine Learning Pipeline.

3.1 Research Framework

The research framework is described. It begins with pre-processing, learning, prediction and evaluation. Figure 1 depicts the entire research strategy for how each facial key point dataset will yield accuracy.

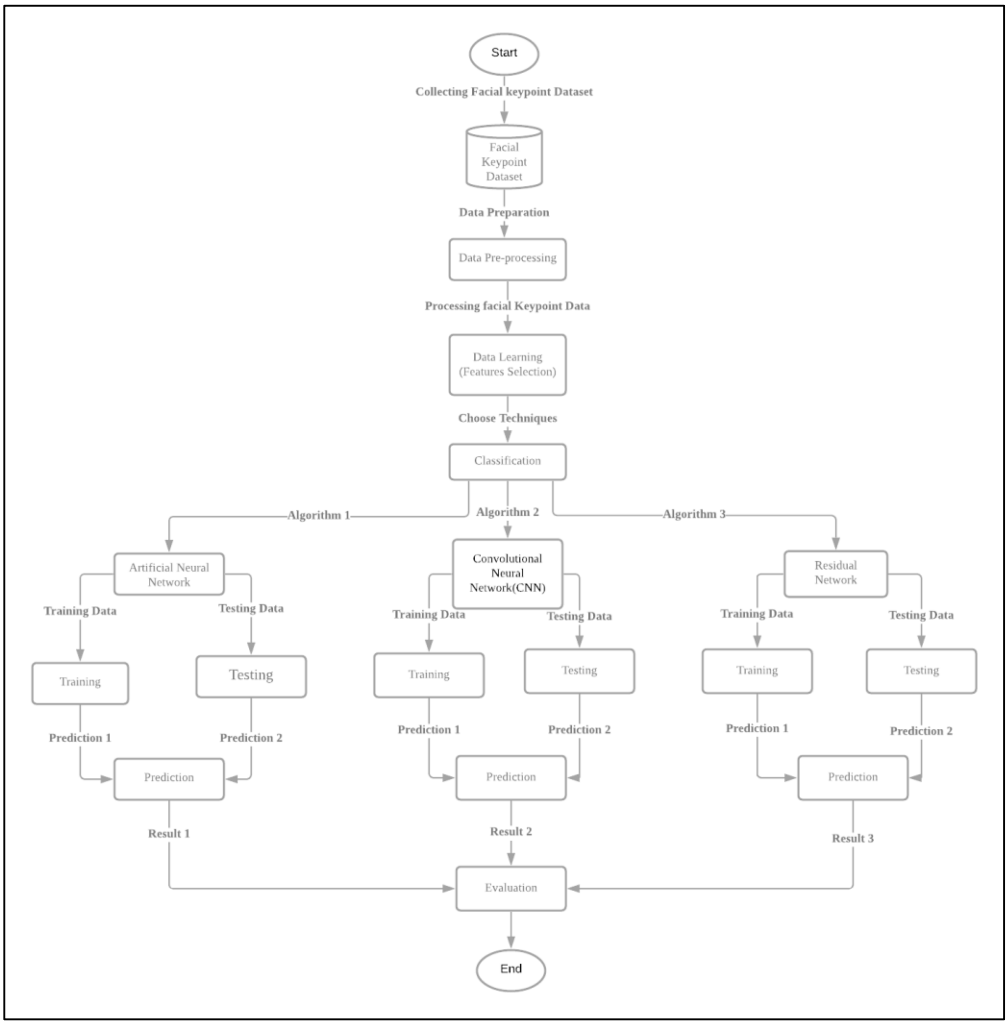

Figure 1 Research framework

3.2 Data Selection

For this phase, data that will be used in Facial detection dataset that was taken from Kaggle - https://www.kaggle.com/c/facial-keypoints-detection/data. This dataset was built by collecting images that consist of facial emotions. Each predicted key point is specified by an (x,y) real-valued pair in the space of pixel indices. There are 15 key points, which represent the following elements of the face:

Left eye center, right eye center, left eye inner corner, left eye outer corner, right eye inner corner, right eye outer corner, left eyebrow inner end, left eyebrow outer end, right eyebrow inner end, right eyebrow outer end, nose tip, mouth left corner, mouth right corner, mouth center top lip, mouth center bottom lip. Left and right here refers to the point of view of the subject.

In some examples, some of the target key point positions are missing (encoded as missing entries in the csv, i.e., with nothing between two commas). The input image is given in the last field of the data files, and consists of a list of pixels (ordered by row), as integers in (0,255). The images are 96x96 pixels.

For Google Teachable Machine, the data was collected raw. It was captured at that time and then it will undergo training and testing phase. There are 5 classes which are happy, sad, neutral, angry and sad. Each class contains 300 samples of data. The total data set for GTM are 1500 samples.

3.3 Data Pre-processing

For algorithm, it will include Convolutional Neural Network (CNN). A Convolutional Neural Network is a form of artificial neural network that is specifically intended to process pixel input and is used in image recognition and processing. CNNs are image processing in artificial intelligence (AI) systems that employ deep learning to do both generative and descriptive tasks. It frequently using machine vision that includes images and video recognition, recommender systems, and natural language processing (NLP).

A neural network is a hardware and/or software system modelled after the way neurons in the human brain work. Traditional neural networks aren't designed for image processing and must be fed images in smaller chunks. CNN's "neurons" are structured more like those in the frontal lobe, the area in humans and other animals responsible for processing visual inputs. Traditional neural networks' piecemeal image processing difficulty is avoided by arranging the layers of neurons in such a way that they span the whole visual field.

A CNN employs a technology similar to a multilayer perceptron that is optimized for low processing requirements. An input layer, an output layer, and a hidden layer with several convolutional layers, pooling layers, fully connected layers, and normalizing layers make up a CNN's layers. The removal of constraints and improvements in image processing efficiency result in a system that is significantly more effective and easier to train for image processing and natural language processing.

3.4 Parameter Selection

Parameter selection is a process choosing the selected parameter for each technique algorithm that will be used during the training of each data. Each of the techniques will be chosen based on the effect of manipulation on a certain parameter. For Artificial Neural Network (ANN), the parameter used are to tune the number of neurons in each hidden layer. There are three layers which are the input layer, where input data are fed into it, the hidden layers, and the output layer, which contains the output value. The Logistic Regression uses the sigmoid function.

$$P(\mathrm{x}) = \frac{1}{1+ e^{-y}}$$

For Convolutional Neural Network (CNN), there are various layer in CNN network. It is input layer, convolutional layer, pooling layer, fully connected layer and output layer. The final difficulty in CNN is the first fully connected layer. This is because the dimensionality of the fully connected layer was not determined. To calculate it, it needs to start with the size of the input image and the size of each convolutional layer was calculated. The parameter that can be tuned is number of epochs. Epochs is a hyperparameter that defines how many times that learning algorithm work through the entire dataset

3.5 Experimental setting

This section describes the experimental setup that will be used in this research study. For this experiment, the algorithm model that are used is that the Artificial Neural Network (ANN), Convolutional Neural network (CNN), and Residual Network (RESNET) as a training and testing model to create this research undefeated manufacture the result. The reason of the use of these models is in training model will be used each of the algorithms to make sure the algorithm can be trained by using crop data set because each algorithm cannot be applied directly because the incompatible of the data and algorithm. That is why the algorithm need to train first before the real test performance.

The validation model is where it will perform the real test to get the output to show the accuracy and performance. The end of the result of each algorithm will validate and evaluate to see the expected outcome of the result whether it meets or not the proposed in this research project.

3.6 Research Activities

There are total of four phases will be used from Pipeline model. Table 4 shows the research activities or milestone in each phase to be conduct and followed during the entire research.

Table 2 Research activities

Phase	Task	Output
Pre-processing	• Collect initial data (acquire the data listed in the project resources for facial emotion classification such as happy, sad, surprise, confuse and angry faces). • Describe data (examine the properties of the acquired data and report on the results). • The data will undergo pre-processing. • The raw data will transform into understandable format.	• Initial data collection report consists of list of the dataset acquired, the method used to acquire it and any problems encountered. • Data description report evaluate whether the data acquired satisfied the relevant requirements. • Data exploration report (including first findings or initial hypothesis and impact on the remainder of the project) • Data quality report (list of the results of the quality verification)
Learning	• Learning algorithm used to process understandable data in order to extract patterns. • There will be dataset that consist of x and y coordinates of 15 facial key points.	• A specific input-output transformation task
Evaluation	• Evaluate the performance of machine learning model. • The number of incorrect predictions counted to determine model prediction accuracy. • The data will be divided into training (80%) and testing (20%).	• The model accuracy prediction determined. • The number of wrong predictions on the test dataset to compute the model's prediction accuracy is counted
Prediction	• The model's performance in determining the outcomes of the test data set was not used for training or cross validation.	• Result of test dataset

4. Results and Discussion

The experiment will start where the dataset has been decided to be used in this research project. The dataset has a feature where each of data has some kind of feature about facial expression that can be used. After downloading from Kaggle.com and securing the dataset, the data need to do some data preparations. It is to make it easy to read by converting dataset into an excel file .CSV. Since the dataset doesn't have any missing value, the data preprocessing step was skipped. Then, jump to the learning algorithm. Learning algorithm is used to process understandable data in order to extract patterns. The goal is to use a system for a specific input-output transformation task. It also to remove noise or outliers or to complete the data if there are some missing values.

There is also the usage of Google Teachable Machine because it is easier to training and testing the data. The number of epochs also can be change directly during the training phase.

In the next phase of the experiment is to apply data that has been processed from before into the selected classification techniques as has been mention earlier. During this phase, the facial key point data will divide into 70% for training and 30% for testing. Sometimes, it might include cross validation dataset as well and then it will be divided into 60% training, 20% validation and 20% testing. These ratios for training and testing will be applied to the same setting for each classification techniques. After all the process is done, every result will be recorded and need to be evaluated either where the accuracy of the data can be more recognize or not.

4.1 Simulation setup

The dataset for Kaggle was tested in Kaggle Notebook. For Google Teachable Machine the data was captured during webcam and undergo training and testing. The splitting ration

Table 3 Data information

No.	Data	Number of Data	Number of the Attribute	Class Label	Column	Row
1.	Facial Recognition dataset (ANN)	8832	5	1	6	8833
2.	Facial Recognition dataset (ResNet)	8832	5	1	6	8833
3.	Google Teachable dataset (CNN)	1500	5	1	6	1501

Table 4 Data split for 70-30

No.	Data	Training (0.70)			Testing (0.30)		
		Kaggle (CNN)	Kaggle (ANN)	Teachable Machine (CNN)	Kaggle (CNN)	Kaggle (ANN)	Teachable Machine (CNN)
1.	Facial emotion recognition	1-7049 (7049)	1-24979 (24979)	1-1050 (1050)	7050-8832 (1783)	24980-35685 (35685)	1051-1500 (450)

4.2 Parameter and Testing Method

A parameter is a model that is used to make a prediction, it refers to the classification approach which is Convolutional Neural Network, Artificial Neural Network, and Residual Network that will be used. Testing method pertain to what the evaluation uses to support the outcome that was obtained from each technique, whether true or false. The method used was by using a confusion matrix that can show the accuracy of the technique, performance, and error of the experiment. Accuracy is calculated by dividing the total quantity of data in the data set by the number of correct predictions. Sensitivity is a calculated all the correct prediction that has divided into a total number correct prediction. Specificity is calculated all the correct prediction that has divided into a total number of incorrect predictions Mostly the best value for accuracy, sensitivity, and specificity were 1.0 or 100%, the worst value was 0.0 or 0%. This equation will be used during the training and testing model to find the accuracy of each algorithm used. The main measure of performance is evaluated in terms of accuracy, precision, and recall from the confusion matrix of classification. The measures are computed by using equations that are described below:

1. Accuracy: The total number of samples correctly classified to the total number of samples classified.

$$Accuracy = \frac{TN + TN}{TP + TN + FN + FP} = \frac{TP + TN}{P + N}$$

2. Recall: The number of samples is classified as positive divided by the total sample in the testing set positive category.

$$Sensitivity = \frac{TP}{TP + FN} = \frac{TP}{P}$$

3. Precision: The number of samples is categorized positively classed correctly divided by total samples are classified as positive samples.

$$Specificity = \frac{TN}{TN + FP} = \frac{TN}{N}$$

4.3 Result

Table 5 The difference between number of epoch and percentage accuracy

Emotion	Number of epochs		
	50	60	85
Happy	82%	84%	97%
Sad	0%	34%	80%
Neutral	100%	100%	100%
Angry	94%	94%	100%
Surprise	100%	100%	100%

5. Conclusion

The objectives of this research are to designing a model that can classify people's emotions based on their face images, to testing a model that train and deploy people emotions and expressions automatically and to evaluate the results that has been tested.

The last objective is to evaluate the result that has been tested based on accuracy rate and precision. It is shows that by increasing the number of epochs, it can increase the accuracy of the emotion detection. From this research, it shows that if the epoch is 50, the percentage of sad was increased from 0 to 80% accuracy. The percentage for happy emotion also increased from 82% to 97%.

The strength of this research is the data was gained raw. So, it is easier to compare with previous research because the other two methodologies were used Kaggle data set. Next, it shows the advanced of Convolutional Neural Network can improve the accuracy result. The limitations of this project are the issue with the dataset, this project can be used to test further different data set and analytically examine the different results. The other limitations are using the parameters, different parameters can be set and used.

This research proposed the new rough set theory and still needs some improvement. In the future, this research is hoped to explore more factors contributing to classify the facial emotion with the other types of algorithms based on the facial key point data set. Besides, another method will be looked for handling the training and testing data in other to improve the performance of each algorithm.

Acknowledgment

The authors would like to thank the Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia for its support.

References

[1] Roslan, S. (2021, September 10). *Kes Bunuh Diri meningkat 143 peratus*. Berita Harian. Retrieved December 28, 2021, from https://www.bharian.com.my/berita/nasional/2021/09/862128/kes-bunuh-diri-meningkat-143-peratus

[2] Mehta, D., Siddiqui, M. F. H., & Javaid, A. Y. (2019, April 21). *Recognition of emotion intensities using Machine Learning Algorithms: A comparative study*. Sensors (Basel, Switzerland). Retrieved December 28, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6514572/

[3] Burns, E., & Burke, J. (2021, March 26). *What is an artificial neural network (ANN)?* SearchEnterpriseAI. Retrieved December 28, 2021, from https://www.techtarget.com/searchenterpriseai/definition/neural-network

[4] Davenport, T., Guha, A., Grewal, D., & Bressgott, T. (2019, October 10). *How artificial intelligence will change the future of Marketing - Journal of the Academy of Marketing Science*. SpringerLink. Retrieved December 28, 2021, from https://link.springer.com/article/10.1007/s11747-019-00696-0

[5] Hassouneh, A., Mutawa, A. M., & Murugappan, M. (2020, June 12). *Development of a real-time emotion recognition system using facial expressions and EEG based on machine learning and deep neural network methods*. Informatics in Medicine Unlocked. Retrieved December 28, 2021, from https://www.sciencedirect.com/science/article/pii/S235291482030201X

[6] Nguyen, B. T., Trinh, M. H., Phan, T. V., & Nguyen, H. D. (1970, January 1). *An efficient real-time emotion detection using camera and facial landmarks: Semantic scholar*. undefined. Retrieved December 28, 2021, from https://www.semanticscholar.org/paper/An-efficient-real-time-emotion-detection-using-and-Nguyen-Trinh/7afc3db34d9e16a70064ee4937b0425045cdb4e8

[7] *Development of a real-time emotion recognition system* ... (n.d.). Retrieved December 28, 2021, from https://www.researchgate.net/publication/342126542_Development_of_a_Real-Time_Emotion_Recognition_System_Using_Facial_Expressions_and_EEG_based_on_machine_learning_and_deep_neural_network_methods/fulltext/5ee397d4458515814a585937/Development-of-a-Real-Time-Emotion-Recognition-System-Using-Facial-Expressions-and-EEG-based-on-machine-learning-and-deep-neural-network-methods.pdf

[8] Fulmer, J., By, Fulmer, J., .tdb-Author-Photo, /* C. C. */.tdb-A.-B., Jenn FulmerJenn Fulmer is a content writer for TechnologyAdvice, & Jenn Fulmer is a content writer for TechnologyAdvice. (2021, September 27). *The value of emotion recognition technology*. IT Business Edge. Retrieved December 28, 2021, from https://www.itbusinessedge.com/business-intelligence/value-emotion-recognition-technology/

[9] *multimodal affective computing to enhance the user* ... (n.d.). Retrieved December 28, 2021, from https://www.researchgate.net/publication/327638354_Multimodal_Affective_Computing_to_Enhance_the_User_Experience_of_Educational_Software_Applications