Synthesis of Compound Facial Expressions Based on Indonesian Sentences Using Multinomial Naïve Bayes Model and Dominance Threshold Equations

Aripin, *Member*, *IAENG*, Hanny Haryanto, and Wisnu Agastya

***Abstract*—This research aimed to synthesize compound facial expressions based on Indonesian sentences using a multinomial naïve Bayes (MNB) model and a dominance threshold equation. A compound facial expression is a combination of several basic emotions. For example, a disappointed facial expression is a combination of the basic emotions of anger and sadness. Compound facial expressions on virtual characters are needed in the animation and gaming industry. In this research, compound facial expressions were formed based on Indonesian sentences. The proposed method uses a combination of an MNB model and a dominance threshold equation, which work in sequence to map the dominant emotion classes. The MNB model is used for emotion classification to obtain the probability value for each class. Then the probability values are used to experimentally determine the dominance threshold value with the dominance threshold equation. The emotion classes that have a probability value greater than the threshold value are chosen as the dominant emotion classes. The combination of two dominant emotional classes forms a compound facial expression. Classification experiments were performed 9 times using a varying ratio between the training data and test data. The experimental result had an accuracy rate of 91.0% with a ratio between the training data and test data of 90 to 10. The compound facial expression system developed in this research was evaluated using a user acceptance test. The number of respondents involved in the evaluation process was 32. The respondents evaluated the system for 3 suitability assessment categories, namely mouth movements, phoneme pronunciation sounds, and facial expressions. The average score for each category was calculated using the mean opinion score (MOS) method. The average user acceptance test score for the 3 assessment categories was 4.21 on a scale of 1 to 5. This shows that the proposed system is capable of producing talking-head animations accompanied by compound facial expressions that look natural and realistic.**

***Index Terms*—Compound expressions, dominance threshold equation, Indonesian sentences, multinomial naïve Bayes**

Manuscript received January 12th, 2021; revised November 4th, 2021. This research was supported by the Ministry of Research, Technology, and Higher Education through a competitive grant program within the Higher Education Excellence Applied Research scheme (Grant No. 312/E4.1/AK.04.PT/2021).

Aripin is a lecturer at the Biomedical Engineering Department of Dian Nuswantoro University, Indonesia (email: arifin@dsn.dinus.ac.id).

Hanny Haryanto is a lecturer at the Informatics Engineering Department of Dian Nuswantoro University, Indonesia (email: hanny.haryanto@dsn.dinus.ac.id).

Wisnu Agastya is a student at the Informatics Engineering Department of Dian Nuswantoro University, Indonesia (email: wisnu.aga@gmail.com).

I. Introduction

There are two types of communication, namely verbal and non-verbal. Verbal communication is direct communication using spoken and written language, while non-verbal communication generally uses body language such as facial expressions, head shaking, hand movements, and actions. Verbal as well as non-verbal communication require facial expressions to emphasize the emotional condition experienced by a person so that communication can take place more effectively [1]. In computer technology, facial expressions can be developed for intelligent virtual characters in animated film productions, games, and virtual speaker animation. Virtual characters need to be able to display realistic and natural facial emotional expressions. Many applications involve realistic and natural expressive virtual characters using English or other foreign languages. However, applications that use Indonesian are still rare. This was one of the challenges in this research.

Indonesian is the official language of the Indonesian nation, which has a large diversity of ethnicities and languages. Indonesian also plays an essential role as the unifying language of the country and has an important position in the international world [2], [3]. The Indonesian language is currently studied in more than 45 countries and the number of visits by foreigners to Indonesia is increasing [4].

Fig. 1. The six basic emotions

The Indonesian sentence structure generally consists of a subject, predicate, object, complement, and description. In general, Indonesian sentence structures are classified based on the number of clauses (ideas in a sentence) into single sentence structures and compound sentence structures [5]. A single sentence is a sentence with one clause, whereas a compound sentence is a sentence that has more than one clause, with each clause separated by a conjunction. In verbal communication using written language, the sentence structure plays an essential role in determining emotion classes.

Emotion is a condition of a person's feelings that can affect behavior, changes in physiology, thinking, and facial expression gestures. A person's emotions are influenced by stimuli from outside and from within a person. Emotions can be transmitted to other people who pay attention to that person's emotional condition. For example, someone experiencing positive emotional conditions such as happiness or joy will make other people who pay attention to them feel happy too. This also applies to negative emotions such as sadness, anger and fear [6].

Emotions expressed in Indonesian sentences can be recognized through the adjectives used [1]. Indonesian language adjectives can express emotions such as anger, disgust, fear, happiness, sadness or surprise. In this research, the emotion classes that were used as reference were the six basic emotion classes distinguished by Ekman & Cordaro [6], i.e. anger, disgust, fear, happiness, sadness, and surprise, as shown in Fig. 1. Anger is characterized by feelings of frustration and disappointment. Disgust is characterized by feelings of resistance and discomfort. Fear is characterized by feelings of anxiety about something that is threatening. Happiness is characterized by feelings of success or satisfaction. Sadness is characterized by feelings of failure or regret. Meanwhile, surprise is characterized by a feeling being shocked by something that comes unexpectedly [7].

A recent study published in the journal *Proceedings of the National Academy of Sciences* identified 15 compound emotions that are a combination of two basic emotions [8], [9], i.e. happy-surprise, sadness-surprise, anger-surprise, fear-surprise, disgust-surprise, happy-disgust, sad-disgust, angry-disgust, disgust-fear, sad-fear, angry-sad, angry-fear, hate, and fascination [10], [11]. Although they may seem similar, these compound emotions do differ from one another. For example, happy-surprise is very different from happy-fear or happy-disgust. In addition to the six basic emotions there are also secondary emotions, i.e. a combination of two or more basic emotions, for example, moved and disappointed [12]. Moved is a combination of happy and sad emotions, while disappointed is a combination of anger and sadness. The development of compound facial expression animation based on Indonesian text has never been conducted before. The present investigation is a preliminary research on this topic.

Mapping of emotion classes from Indonesian sentences can be done using classification algorithms and applications from the natural language processing (NLP) field. In this paper, we propose the mapping of emotions based on Indonesian sentences using a multinomial naïve Bayes (MNB) classification algorithm. MNB is a probabilistic model based on the naïve Bayes algorithm. This model performs data processing using a multinomial distribution by measuring the level of intensity of a category. In emotion mapping, the frequency of adjectives represents emotion intensity.

The results of the text classification process using the MNB classifier produces only one emotion class with the highest emotional intensity. However, sentences that have a compound structure usually represent more than one emotion class. Therefore, a dominance threshold equation is used to complement the MNB classifier to identify the compound emotion classes. First, the MNB classifier is used to calculate the probability value of each emotion class. Next, the dominance threshold value is calculated using the dominance threshold equation based on the probability value of each emotion class. The probability value of each class is

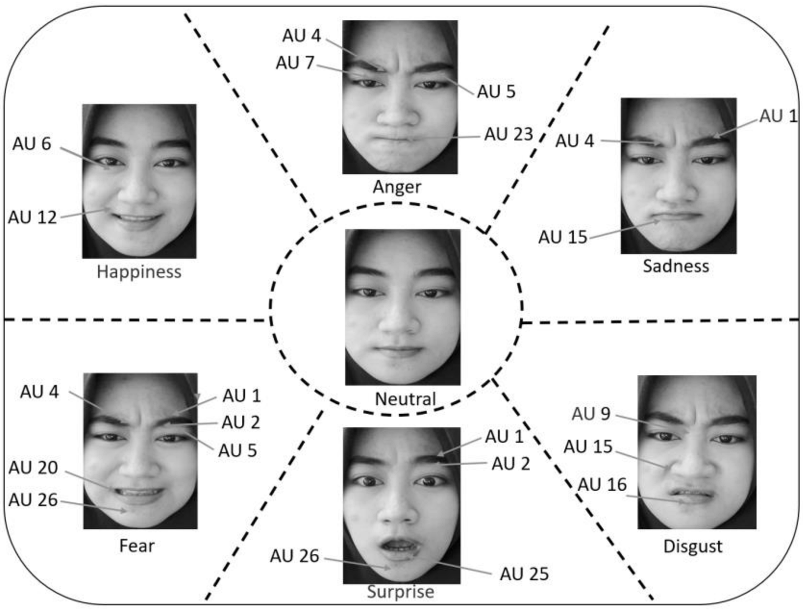

Fig. 2. Action units for each basic emotion

compared with the threshold value. If the probability value of an emotion class is greater than or equal to the threshold value, then it is categorized as dominant.

The facial expression of a person's emotions can be recognized through the Facial Action Coding System (FACS) developed by Ekman & Friesen [15]. This is a facial expression coding system based on facial muscle movements. Facial expression elements produced by facial muscle movements are called action units (AUs) [13], [14]. There are three categories of AUs, i.e. main action units, head movement action units, and eye movement action units [15]. Each basic emotion consists of several AUs, as shown in Fig. 2. Meanwhile, a compound facial expression combines several AUs, each related to a basic emotion.

This research aimed to develop compound facial expression animation based on Indonesian sentence input using a multinomial naïve Bayes (MNB) model. This paper is divided into several sections: 1) introduction, explaining the background of the research problem; 2) related work, explaining previous related studies; 3) methodology, explaining the research steps from data acquisition to compound facial expression animation; 4) experimental results, explaining the experimental testing of the proposed compound facial expression system; and 5) evaluation, explaining the validation of the classification process.

II. Related Work

Some of the related research that has been done previously is described in this section. The research by Wongso et al. [16] on text classification of news articles in Indonesian used a multinomial naïve Bayes model for text classification and TF-IDF for feature extraction. The precision and recall values produced were higher compared to using feature extraction and other methods. A multinomial naïve Bayes algorithm was also used to classify text data from Twitter for disease mapping in Indonesia [17]. The method was able to reveal diseases and their location with highly accurate map visualization. The research by Yovellia Londo et al. [18] on Indonesian news article text classification explored machine learning algorithms such as SVM, MNB, and Decision Tree to classify news articles. The results showed that the MNB algorithm achieved the highest F1 score.

The multinomial naïve Bayes approach is capable of classifying Indonesian text with excellent performance. In this paper, we propose a dominance threshold equation to select dominant emotion classes from the probability value of each emotion class. The combination of the multinomial naïve Bayes model and the dominance threshold equation was used to map Indonesian sentences into compound facial expressions. The addition of the dominance threshold equation to the classification process with the MNB model is a novelty of this research.

III. Methodology

The steps of the proposed method are presented in Fig. 3. The research step sequence consisted of data acquisition and oversampling, text preprocessing, classification, determining the threshold value by using the dominance threshold equation, selecting the dominant classes, and animating virtual characters based on the dominant emotion classes.

A. Data Acquisition and Oversampling

The dataset used in this research consisted of 1306 sentences. These sentences were collected from various sources, such as books, magazines, newspapers, and digital media. The dataset contained single sentences and compound sentences. Each sentence was labeled based on emotion classification. Linguists and psychologists validated the dataset to ensure that the data corresponded to emotion classes. The number of adjectives representing a particular class of emotions was identified to ensure the balance of the number of adjectives. The result of the identification of the number of adjectives is presented in Table I. The difference in the number of adjectives for each emotion class was too large, which could have an impact on the validity of the classification results. Oversampling was used to overcome this problem, so that the amount of data was balanced, as shown in Table II.

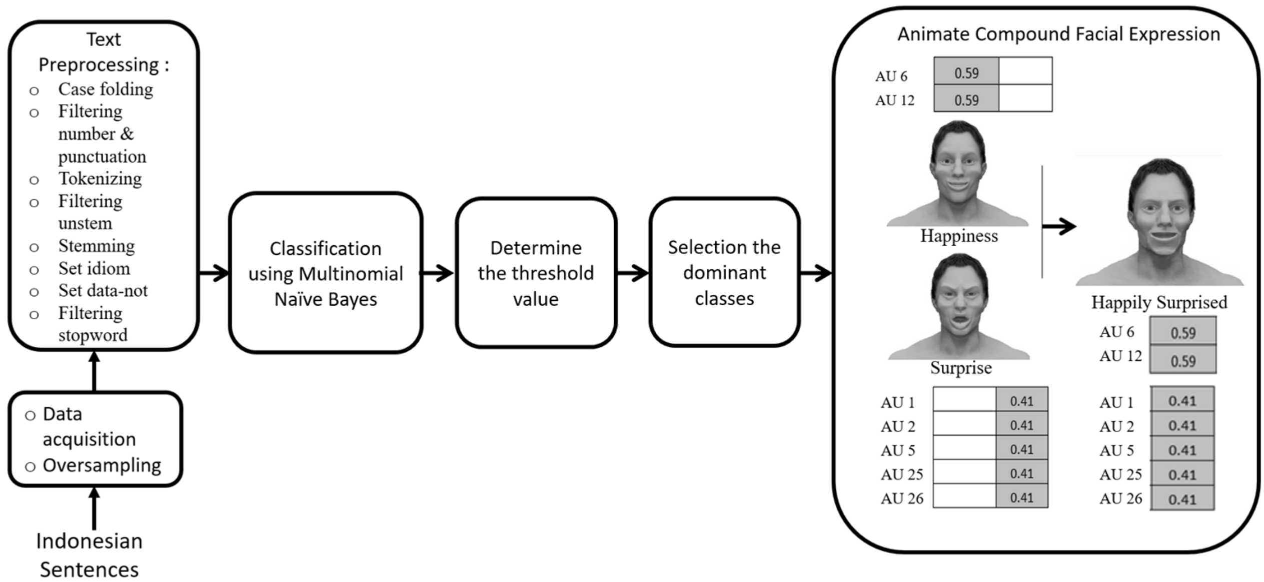

Fig. 3. The Proposed Method

TABLE I
NUMBER OF ADJECTIVES FOR EACH EMOTION CLASS

Emotion class	Number of adjectives
Anger	276
Disgust	98
Fear	137
Happiness	326
Sadness	367
Surprise	102

TABLE II
NUMBER OF ADJECTIVES FOR EACH EMOTION CLASS AFTER OVERSAMPLING

Emotion class	Number of adjectives
Anger	365
Disgust	364
Fear	362
Happiness	365
Sadness	367
Surprise	364

B. Text Preprocessing

Each sentence of the dataset was processed in the text preprocessing step to extract data features. This process includes case folding, number and punctuation removal, tokenization, non-stemming, stemming, set idiom, set data-not, and stopword removal.

The first step of preprocessing is case folding, which changes all letters to lowercase. The next step is number and punctuation removal, which deletes data that has no information related to class selection. Tokenization splits compound words. The next step is non-stemming, i.e. selecting words that must not enter the stemming step. The words chosen in this step are not converted into root words as this can change their meaning. For example, the word *meninggal* cannot be changed into the root word *tinggal*, because *meninggal* and *tinggal* have different meanings. The stemming step changes words into their roots by removing the affixes. The next step is set idiom, which compiles the root words in an idiom word list. For example, the word *tanggung* is combined with the word *jawab* to become *tanggung jawab*. The set-idiom step avoids words that are included in the idiom but are still separate. Meanwhile, the set data-not step keeps separate words together to preserve the meaning of an expression. For example, the expression *tidak suka* has a different meaning from the word *suka*. Finally, the stopword removal step removes conjunctions such as *yang*, *di*, *dan*, and *pada*.

C. Multinomial Naïve Bayes Classification

Multinomial naïve Bayes (MNB) classifiers and are often used in text classification. They are very suitable for classifying data with discrete features, for example counting the number of words in a document [19], and have excellent performance [20]. In this research, a MNB classifier was used to count the numbers of certain words in Indonesian sentences [21], [22]. The classifier determines classes by the words that appear in a document and the number of occurrences of each word. The first step is to calculate the probability of a document d being in class c using (1).

$$P(c|d) \propto P(c)\prod_{k=1}^{n} P(t_k|c) \quad (1)$$

where $P(c|d)$ is the probability of a document d being in class c, $P(c)$ is the prior probability of the document being in class c, $\{t_1, t_2, t_3, \dots, t_n\}$ are the tokens in the document that are part of the vocabulary, and $P(t_k|c)$ is the conditional probability of the term t_k being in the document in class c.

Document classification aims to assign the best class to a document by finding the maximum a posteriori (C_{map}) class using (2). The result of the calculation is the conditional probability multiplied, which can cause floating-point underflow. Therefore, the calculation process needs to be improved by summing the log probability. The class with the highest log probability is the class with the highest probability from the document. Calculating the log probability is done using Equation (4). P is denoted as $\hat{P}$ because the values of $P(c|d)$ and $P(t_k|c)$ are still unknown; they will be calculated during the training process.

$$C_{map} = \arg\max_{C\epsilon\{C_l|C_S\}} \hat{P}(c|d) \quad (2)$$

$$= \arg\max_{C\epsilon\{C_l|C_s\}} \hat{P}(c_j)\prod_{k=1}^{n} \hat{P}(t_k|c_j) \quad (3)$$

$$= \arg\max_{C\epsilon\{C_l|C_s\}} \left[\log \hat{P}(c) + \sum_{1\le k\le n} \log \hat{P}(t_k|c)\right] \quad (4)$$

$\hat{P}(c)$ and $\hat{P}(t_k|c)$ are obtained by calculating the maximum likelihood, which is the relative frequency of the parameter. $\hat{P}(c)$ is calculated using (5), while $\hat{P}(t_k|c)$ is calculated using (6).

$$\hat{P}(c) = \frac{N_c}{N} \quad (5)$$

where $P(c)$ is the prior probability of a document being in class c, N_c is the number of documents in class c, and N is the total number of documents.

$$\hat{P}(t|c) = \frac{T_{ct}}{\sum_{t' \in V} T_{ct'}} \quad (6)$$

where $\hat{P}(t|c)$ is the conditional probability of the term t being in the documents in class c, T_{ct} is the number of occurrences of term t in documents in class c, and $\sum_{t' \in V} T_{ct}$ is the frequency of all terms in class c.

The maximum likelihood calculation can produce a value of 0 caused by a word in the class that is not visible in the training data. This causes the calculation of $P(c|d)$ to be 0. Therefore, the maximum likelihood calculation is done using the Laplace smoothing technique so that it becomes (7).

$$\hat{P}(t|c) = \frac{T_{ct}+1}{\sum_{t' \in V}(T_{ct'}+1)} = \frac{T_{ct}+1}{(\sum_{t' \in V} T_{ct'})+B'} \quad (7)$$

where B is the number of all terms in the vocabulary. Meanwhile, the MNB formula that uses term frequency-inverse document frequency (TF-IDF) weighting is expressed in (8).

$$\hat{P}(t|c) = \frac{W_{ct}+1}{(\sum_{w' \in V} W_{ct'})+B'} \tag{8}$$

where W_{ct} is the weight of TF-IDF term t in documents in class c, $\sum_{w' \in V} W_{ct'}$ is the number of TF-IDF weights for all terms in class c, and B' is the IDF for all terms in the vocabulary.

The results of the MNB model are probability values. The probability values are converted into a percentage value to get a universal value. This is also intended to avoid different probability values for each sentence. The formula used is (9).

$$Percentage_i = \hat{P}\left(t_i \middle| c_j\right) x100 \tag{9}$$

where $Percentage_i$ is the percentage of the i^{th} emotion class, and $\hat{P}\left(t_i \middle| c_j\right)$ is the probability of the i^{th} emotion class.

TABLE III
CONFUSION MATRIX

		Actual Values	
Predicted Values		Positive (1)	Negative (0)
	Positive (1)	TP (true positive)	FP (False Positive)
	Negative (0)	FN (False Negative)	TN (True Negative)

$$Precision = \frac{TP}{(TP+FP)} \tag{10}$$

$$Recall = \frac{TP}{(TP+FN)} \tag{11}$$

$$Accuray = \frac{(TP+TN)}{(TP+FP+FN+TN)} \tag{12}$$

$$F1\ Score = 2x\ \frac{Recall\ x\ Precision}{Recall + Precision} \tag{13}$$

The confusion matrix contains information about the classifier's actual and predicted values [23] as shown in Table III. The measurement of classification performance is accuracy, where the amount of data classified correctly is divided by the total amount of data. Precision is the ratio between a true positive prediction and the overall positive prediction result. Recall is the ratio between the true positive predictions and the overall true positive data. Meanwhile, the F1 score is a comparison of the average precision and recall. Equations (10), (11), (12), and (13) are used to calculate precision, recall, accuracy, and F1 score, respectively.

D. Determining the Dominance Threshold Value

The percentage value of each emotion class is used to calculate the threshold value with (14).

$$D = min(p) + \beta \tag{14}$$

where *D* is the dominance threshold value, *min(p)* is the smallest probability percentage value, and β is the optimal value determined based on the experimental results. The dominance threshold value generated from the calculation using (14) is the value in percentage form.

E. Selection of the Dominant Classes

The dominant emotion classes are the emotion classes that have a percentage value greater than the threshold value. A compound facial expression is obtained by combining the dominant emotional classes. The percentage value for each class of dominant emotions is normalized using (15) so that the total percentage of dominant emotion classes for each sentence is 100%.

$$norm(z)_i = \frac{z_i}{\sum_{j=0}^{K} z_j} \tag{15}$$

where z_i is the i^{th} data from the input data, $\sum_{j=0}^{K} z_j$ is the total number of input data, and *K* is the length of the input data.

F. Compound Facial Expression Animation

Compound facial expression animation is done in two steps, namely determining the associated AUs and combining the associated AUs.

TABLE IV
MAIN ACTION UNITS

AU#	FACS name	AU#	FACS name
0	Neutral face	15	Lip corner depressor
1	Inner brow raiser	16	Lower lip depressor
2	Outer brow raiser	17	Chin raiser
4	Brow lowerer	18	Lip pucker
5	Upper lid raiser	19	Tongue show
6	Cheek raiser	20	Lip Stretcher
7	Lip tightener	21	Neck tightener
8	Lips toward each other	22	Lip funneler
9	Nose wrinkler	23	Lip tightener
10	Upper lip raiser	24	Lip pressor
11	Nasolabial deepener	25	Lips part
12	Lip corner puller	26	Jaw drop
13	Sharp lip puller	27	Mouth stretch
14	Dimpler	28	Lip suck

Main action units are AU categories that are closely related to compound facial expressions. They consist of 28 AUs [24], which are marked with serial numbers as shown in Table IV. Each basic emotion is composed of several AUs as shown in Table V.

Each AU represents a facial muscle movement that is used to form facial expressions. For example, AU4 is the contraction of two muscles resulting in drooping of the brow (with emphasis on the inner side). This AU is used to express sadness, fear, and anger.

Compound facial expressions are formed by combining AUs related to each of the basic emotion classes. Several patterns of combining basic emotions into compound facial expressions [25] are shown in Table VI.

TABLE V
ACTION UNITS FOR EACH EMOTION CLASS

Basic emotion	Associated AUs
Happiness	AU6, AU12
Surprise	AU1, AU2, AU5, AU25, AU26
Disgust	AU9, AU15, AU16
Fear	AU1, AU2, AU4, AU5, AU7, AU20, AU26
Anger	AU4, AU5, AU7, AU23
Sadness	AU1, AU4, AU15

TABLE VI
PATTERNS OF COMBINING BASIC EMOTION CLASSES

Basic emotion#1	Basic emotion#2	AUs of compound facial expression
Happiness	Surprise	Happily Surprised: AU1, AU2, AU5, AU12, AU25, AU26
Happiness	Disgusted	Happily Disgusted: AU6, AU12, AU9, AU15
Surprise	Fear	Fearfully Surprised: AU1, AU2, AU5, AU20, AU25, AU26
Surprise	Disgust	Disgustedly Surprised: AU1, AU2, AU5, AU 9, AU15, AU16
Disgust	Fear	Fearfully Disgusted: AU1, AU2, AU4, AU5, AU9, AU15, AU20, AU26
Happiness	Sad	Moved: AU1, AU4, AU6, AU12, AU15
Sadness	Anger	Disappointed: AU1, AU4, AU5, AU15, AU7, AU23
Anger	Fear	Anger and Fear: AU1, AU2, AU4, AU5, AU7, AU15, AU20, AU23

IV. EXPERIMENTAL RESULTS

We conducted 9 experiments using varying ratios between the training data and test data. For each experiment, the precision, recall, and accuracy values were calculated, as presented in Table VII. These experiments were aimed at finding the optimal ratio between the training data and test data for achieving the best classification performance.

Next, we describe in detail the process from determining the dominant emotion classes to compound facial expression animation. The Indonesian sentences used as sample test data are shown in Table VIII. Each sentence was tested on the system to calculate the probability value. Each probability value was converted to a percentage using (9). The percentages resulting from of the conversion of the probability values are shown in Table IX.

The dominance threshold was calculated using (14). The value of β for each sentence differed depending on the results of the experiment for each sentence. In this case, we use sentence#1 as an example for calculating the dominance threshold value. This process begins by selecting the emotion class with the smallest probability value, which in this case was 5.5% for the happiness class. This value was used as the value for the *min(p)* parameter.

TABLE VII
PRECISION, RECALL, AND ACCURACY

Training set	Testing set	Precision	Recall	Accuracy
10%	90%	65%	57%	56%
20%	80%	72%	70%	70%
30%	70%	79%	78%	77%
40%	60%	81%	80%	77%
50%	50%	82%	80%	80%
60%	40%	85%	83%	83%
70%	30%	86%	85%	85%
80%	20%	89%	89%	89%
90%	10%	91%	90%	91%

TABLE VIII
SAMPLE OF INDONESIAN SENTENCES IN TEST SET

Sentence#	Indonesian sentences
1	*kurang ajar siapa yang berani melakukan ini, apa jangan-jangan itu hantu ya*
2	*aku mendapat kado dari ayah, namun betapa terkejutnya aku ternyata ini adalah sepatu yang selama ini aku inginkan*
3	*akhirnya kegiatan jelajah malam kali ini selesai, namum ketika aku hendak ke toilet aku dikejutkan ular yang tiba-tiba di depanku akupun lari terbirit birit*
4	*Budi merasa sangat murka karena uangnya dicopet dan dia masih harus memikirkan untuk mengembalikan uang tersebut*
5	*sejak kecil dia sudah menjadi anak yatim piatu sehingga diasuh pamannya, namun dia sekarang hidup bahagia karena menjadi orang kaya dan bergelimang harta*

TABLE IX
RESULT OF CONVERTING PROBABILITY VALUES TO PERCENTAGE VALUES

Sentence#	Happiness	Sad	Anger	Fear	Disgust	Surprise
1	5.5%	13.6%	26.6%	34.0%	13.3%	7.0%
2	18.4%	7.9%	6.4%	9.4%	6.3%	51.5%
3	6.4%	4.8%	10.0%	30.3%	9.2%	39.2%
4	9.1%	32.3%	23.3%	11.6%	11.1%	11.7%
5	25.3%	30.2%	13.7%	7.7%	7.8%	15.2%

The *min(p)* value was used to conduct experiments on β values from 0% to 10%. In each experiment, the *min(p)* value was added to the β value to calculate the dominance threshold value. The changes in the values of the variables β and D in each experiment are shown in Table X. From the experiment it can be seen that the optimal dominant emotion classes were obtained with $\beta = 9\%$ and dominance threshold = 1.5%. Therefore, the value used as the threshold for sentence#1 was 14.5%. These steps were done in the same way for sentence#2 to sentence#5 to obtain the experimental results.

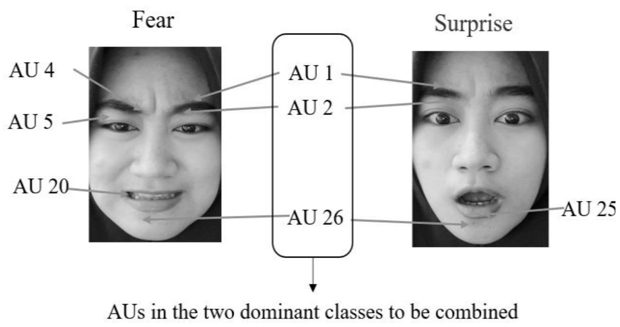

Fig. 4. Combining AUs from dominant emotion classes

The formation of compound facial expressions was done by combining the AUs related to each basic emotion. The result of this combination formed a new emotion class, which was a derived emotion class. Some of these derived emotion classes are representations of a person's general expression. For example, the emotional expression of being disappointed is a combination of the basic emotions of anger and sadness.

TABLE X
EXPERIMENTS TO DETERMINE THE VALUE OF *D*

Min(p)	β	D	Dominant emotion classes
5.5%	0%	5.5%	Sadness, Anger, Fear, Disgust, Surprise
5.5%	1%	6.5%	Sadness, Anger, Fear, Disgust, Surprise
5.5%	2%	7.5%	Sadness, Anger, Fear, Disgust
5.5%	3%	8.5%	Sadness, Anger, Fear, Disgust
5.5%	4%	9.5%	Sadness, Anger, Fear, Disgust
5.5%	5%	10.5%	Sadness, Anger, Fear, Disgust
5.5%	6%	11.5%	Sadness, Anger, Fear, Disgust
5.5%	7%	12.5%	Sadness, Anger, Fear, Disgust
5.5%	8%	13.5%	Sadness, Anger, Fear
5.5%	9%	14.5%	Anger, Fear
5.5%	10%	15.5%	Anger, Fear

TABLE XI
DOMINANT EMOTION CLASSES OF EACH SENTENCE

Sentence #	Dominance threshold value	Dominant emotion classes	Percentage value after normalization
1	15.5%	Anger (26.6%) Fear (34.0%)	Anger (44%) Fear (56%)
2	10.3%	Happiness (18.4%) Surprise (51.5%)	Happiness (26%) Surprise (74%)
3	10.8%	Fear (30.3%) Surprise (39.2%)	Fear (44%) Surprise (56%)
4	12.1%	Sad (32.3%) Anger (23.3%)	Sad (58%) Anger (42%)
5	15.7%	Happiness (25.3%) Sadness (30.2%)	Happiness (46%) Sadness (54%)

Finally, the compound facial expressions derived from each sentence were visualized in 3D virtual characters. The synthesis of compound facial expressions was based on the probability value of each basic emotion class. The FACS of each basic emotion class is composed of several related AUs, as listed in Table V. In developing a compound facial expression animation, each AU is implemented into a blend shape animation with a value ranging between 0 and 1 [26]. The percentage value of each dominant emotion class is used as the weight value for each AU. The resulting compound facial expression animation combines the weight values of the AUs for the two dominant emotion classes [27].

$$AU_{i\in c_j} = {}^{Norm(Percentage_i \geq D)}/_{100} \qquad (16)$$

where AU_i is the i^{th} AU value, D is the dominance threshold value, and $norm(Percentage_i)$ is the percentage value after normalization.

The AUs to be combined were found by finding the two emotional classes with the highest probability values. The process of combining the AUs of the dominant emotional classes is illustrated in Fig. 4. Equation (16) was used to determine each AU value for synthesizing the compound facial expressions. In this research, a compound facial expression system based on Indonesian sentence input was developed, as shown in Fig. 5. Table XI shows the results of the formation of combined facial expressions based on the probability values of the two dominant emotion classes.

the compound facial expression is animated using the morph animation technique. This technique manipulates keyframes by interpolating the entire mesh by creating polygonal surfaces in 3D facial animation [13]. The interpolation method can be calculated using two values ($value_1$ and $value_2$) and the interpolation coefficient α, as expressed in (17).

$$Value = \alpha\,(value_1) + (1.0 - \alpha)\,(value_2) \qquad (17)$$

$$Where\ 0.0 < \alpha < 1.0$$

Based on the experimental results, the combined MNB model + dominance threshold equation could successfully identify two dominant emotion classes. The classification process using only the MNB model is only able to recognize the class with the highest probability value. A comparison of the classification results between the combined MNB model + dominance threshold equation and only the MNB model is shown in Table XII. It shows that the combination of MNB model + dominance threshold equation succeeded in identifying two dominant emotional classes as the basis for the formation of compound facial expressions. The resulting compound facial expressions could represent the content of expressions of anger and fear simultaneously according to the meaning contained in the sentence.

TABLE XII
COMPARISON OF CLASSIFICATION RESULTS USING ONLY MNB MODEL AND COMBINED MNB MODEL + DOMINANCE THRESHOLD EQUATION

Sentence#1	Emotion classes	
	MNB	MNB + dominance threshold equation
kurang ajar siapa yang berani melakukan ini, apa jangan-jangan itu hantu ya	Anger (54.43%)	Anger (44%), Fear (56%)

We evaluated the compound facial expression animations produced by the results of this research. The compound facial expression animations were compared with photos of a person displaying certain compound expressions, as shown in Table XIII. The results of the evaluation showed that there were notable similarities between the compound facial expression animations and the photos of a person's compound facial expressions.

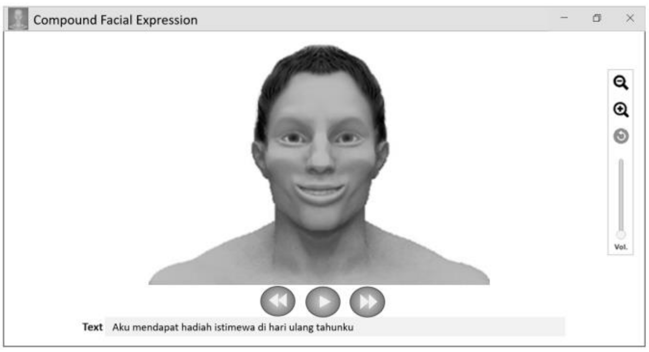

Fig. 5. Compound facial expression system.

TABLE XIII
THE ANIMATION OF COMPOUND FACIAL EXPRESSIONS

Sentence#	*The value of each AU for each dominant emotion class*	*Animation of the compound facial expression*	*Photo of a person's compound expression*	*Type of the compound facial expression*
1	Anger (44%) : 0.44 Fear (56%) : 0.56			Anger and Fear
2	Happiness (26%) : 0.26 Surprise (74%) : 0.74			Happily surprised
3	Fear (44%) : 0.44 Surprise (56%) : 0.56			Fearfully Surprised
4	Sadness (58%) : 0.58 Anger (42%) : 0.42			Disappointed
5	Happiness (46%) : 0.46 Sadness (54%) : 0.54			Moved

In the final step, we tested the compound facial expression system by conducting a user acceptance test. This test involved 32 respondents who used 3 suitability assessment categories, namely lip movements during phoneme pronunciation, phoneme pronunciation sounds, and compound facial expressions. The respondents assessed the system online as shown in Fig. 5 by inputting 5 Indonesian sentences in Table VIII.

$$MOS = \sum_{i=1}^{n} \frac{x(i).k}{N} \quad (18)$$

where $x(i)$ is the sample value of *i*, *k* is the weight, and *N* is the respondent number.

The respondents rated the three categories for each sentence by giving a score of 5 (excellent), 4 (good), 3 (adequate), 2 (bad), and 1 (very bad). A recapitulation of the assessments by the respondents for each category is presented in Tables XIV, XV, and XVI. The average respondent assessment for each category was calculated using the mean opinion score (MOS) method using (18) [28].

TABLE XIV
RECAPITULATION OF RESPONDENT RATINGS FOR THE SUITABILITY ASSESSMENT CATEGORY OF MOUTH MOVEMENTS

Sentence#	Very Bad	Bad	Adequate	Good	Excellent
1	0	0	5	7	20
2	0	1	7	10	14
3	0	0	4	7	21
4	0	0	3	14	14
5	0	0	9	10	13

TABLE XV
RECAPITULATION OF RESPONDENT RATINGS FOR THE SUITABILITY ASSESSMENT CATEGORY OF PHONEME PRONUNCIATION SOUNDS

Sentence#	Very Bad	Bad	Adequate	Good	Excellent
1	0	0	5	5	22
2	0	0	6	11	15
3	0	0	9	7	16
4	0	0	2	14	16
5	0	0	7	10	15

TABLE XVI
RECAPITULATION OF RESPONDENT RATINGS FOR THE SUITABILITY ASSESSMENT CATEGORY OF FACIAL EXPRESSIONS

Sentence#	Very Bad	Bad	Adequate	Good	Excellent
1	0	2	5	8	17
2	0	1	13	8	10
3	0	0	11	7	14
4	0	3	5	14	10
5	0	2	10	12	8

The MOS value for the mouth movements category was 4.30, while the MOS scores for the phoneme pronunciation sounds category and for the compound facial expressions category were 4.34 and 3.99, respectively. The average MOS value for the three categories was 4.21. This shows that the proposed system was capable of producing talking-head animations accompanied by compound facial expression visualization with good results, which means that the compound facial expressions produced by the proposed system looked natural and realistic.

V. CONCLUSION AND DISCUSSION

Based on the experimental results it can be concluded that the proposed MNB model could classify Indonesian sentences so that a probability value for each emotion class was obtained with an accuracy rate of 91.0%. Meanwhile, the dominance threshold equation could be used to select two dominant emotion classes based on a threshold value. The dominant classes are used by the proposed system as the basis for forming compound facial expressions. Compound facial expressions are expressions commonly experienced by a person when facing complex emotional situations.

The animations of compound facial expressions generated from this research were the result of a combination of AUs related to each basic emotion class. The value of each AU was obtained from the probability value of each dominant emotion class after normalization. Based on the evaluation results, the compound facial expression animations succeeded in resembling a person’s compound facial expressions.

Based on the results of the user acceptance test involving 32 respondents, it was shown that the average MOS score for the 3 categories was 4.21. This means that the compound facial expressions produced by the proposed system looked natural and realistic.

In future research, the animation of compound facial expressions can be further developed to become more natural by involving muscle wrinkles and bones in the face. Meanwhile, a classification process that takes into account the semantics of the sentence (not only taking into account the occurrence frequency of adjectives) can be investigated. The relationship and sequence of words in a sentence greatly determines the semantics of the sentence. Thus, the synthesis of Indonesian compound facial expressions is expected to produce more natural and realistic results.

Compound facial expression animation can have a wide impact on various fields, such as the development of animated films and games, virtual speaker animation, human-computer interaction applications, facial expression analysis, and so on. In the field of facial expression analysis, compound facial expressions can improve the diagnosis of psychological disorders or stress disorders.

REFERENCES

[1] J. E. Prawitasari, “Recognizing Emotions through Non-verbal Communication,” *Bul. Psikol,* vol. 3, no. 1, pp27-43, 2016.

[2] Park Jae Hyun, “Potentials and Challenges of Indonesian Language Towards an International Language (translated from:,*Potensi dan Tantangan Bahasa Indonesia Menuju Bahasa Internasional*),” Journal of Sociotechnology, vol. 14, no. 1, pp12-20, 2015.

[3] Sugono, “Indonesian to International Language (translated from: *Bahasa Indonesia Menuju Bahasa Internasional*),” *Fon: Jurnal Pendidikan Bahasa dan Sastra Indonesia*, vol. 11, no. 2, pp1-9, 2017.

[4] Kadek Kevin Reynaldi Nugraha and Made Heny Urmila Dewi, “The Influence of Tourist Visits and Tourist Attraction Income on Poverty in Bali Province (translated from: *Pengaruh Kunjungan Wisatawan dan Pendapatan Daya Tarik Wisata Terhadap Kemiskinan di Provinsi Bali*),” E-Jurnal Ekonomi Pembanagunan Universitas Udayana, vol. 9, no. 3, pp503-529, 2020.

[5] S. S. T. W. Sasangka, “Series of Indonesian Language Education: Sentences (translated from: *Seri Penyuluhan Bahasa Indonesia: Kalimat),” Pusat Pembinaan dan Pemasyarakatan; Badan Pengembangan dan Pembinaan Bahasa; Kementrian Pendidikan dan Kebudayaan, Jakarta*, vol. 1, pp43-52, 2015.

[6] P. Ekman and D. Cordaro, “What is Meant by Calling Emotions Basic,” *Emot. Rev*, vol. 3, no. 4, pp364-370, 2011.

[7] Eun-Hye Jang, Sangwon Byun, Mi-Sook Park, and Jin-Hun Sohn, “Reliability of Physiological Responses Induced by Basic Emotions: A Pilot Study,” Journal of Physiological Anthropology, article number: 15, pp1-12, 2019.

[8] S. Du, Y. Tao and A. M. Martinez, “Compound Facial Expressions of Emotion,” In: *Proceedings of the National Academy of Sciences*, pp. 1454-1462, 2014.

[9] Hongfei Li and Qing Li, “End-to-End Training for Compound Expression Recognition,” Sensors, vol. 20, issue 17, pp1-25, 2020.

[10] Jason Haberman, Pegan Lee, and David Whitney, “Mixed emotions: Sensitivity to facial variance in a crowd of faces,” *Journal of Vision*, vol. 15, issue 4, pp1-11, 2015.

[11] Shichuan Du and Aleix M, Martinez, “Compound facial expressions of emotions: from basic research to clinical applications,” Dialogues in Clinical Neuroscience, vol. 17, no. 4, pp443-455, 2015.

[12] A. Nur Rohman, R. Dwi Handayani, and R. Y. Dwi, “Social Media Emotion Detection Using Term Frequency-Inverse Document Frequency,” *CSRID Journal*, vol. 11, no. 3, pp140-148, 2019.

[13] Ahmad Hoirul Basori and Hani Moaiteq Abdullah AlJahdali, “Emotional Facial Expression Based on Action Units and Facial Muscle,” *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 6, no. 5, pp2478-2487, 2016.

[14] Shangfei Wang, Longfei Hao, and Qiang Ji, “Facial Action Unit Recognition and Intensity Estimation Enhanced Through Label Dependencies,” *IEEE Transactions on Image Processing*, vol. 28, issue 3, pp1428-1442, 2019.

[15] P. Ekman, and W. V. Friesen, “Facial Action Coding System: A Technique for the Measurement of Facial Movement,” *Consulting Psychologies Press*, Palto Alto, CA, 1978.

[16] R. Wongso, F. A. Luwinda, B. C. Trisnajaya, O. Rusli, and Rudy, “News Article Text Classification in Indonesian Language,” In: *Procedia Computer Science*, Bali, Indonesia, pp137-143, 2017.

[17] R. Ranovan, A. Doewes, and R. Saptono, “Twitter Data Classification using Multinomial Naive Bayes for Tropical Diseases Mapping in Indonesia,” *J. Telecommun. Electron. Comput. Eng*, vol. 10, no. 2–4, pp155–159, 2018.

[18] G. L. Yovellia Londo, D. H. Kartawijaya, H. T. Ivariyani, Y. S. P. W.P., A. P. Muhammad Rafi, and D. Ariyandi, “A Study of Text Classification for Indonesian News Article,” In: *2019 International Conference of Artificial Intelligence and Information Technology (ICAIIT)*, pp205–208, 2019.

[19] Muhammad Abbas, Kamran Ali Memon, Abdul Aleem Jamali, Saleemullah Memon, and Anees Ahmed, “Multinomial Naïve Bayes Classification Model for Sentiment Analysis,” *IJCSNS International Journal of Computer Science and Network Security*, vol. 19, no. 3 pp62-67, March 2019.

[20] Shivam Bachhety, Shubham Dhingra, Rachna Jain, and Nikita Jain, "Improved Multinomial Naïve Bayes Approach for Sentiment Analysis on Social Media," *International Journal of Information Systems & Management Science*, vol. 1, no. 1, 2018.

[21] I. H. Witten, F. Eibe, and M. A. Hall, "Data Mining: Practical Machine Learning Tools and Techniques," vol. 3, USA: Elsevier, 2011.

[22] A. S. Thanuja Nishadi, "Text Analysis: Naïve Bayes Algorithm using Python JupyterLab," *International Journal of Scientific and Research Publication*, vol. 9, issue 11, pp126-133, 2019.

[23] B. Liu, "Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data," Second Edition, New York: Spinger, 2011.

[24] Huaixin Chen and Daocai Fu, "An Improved Naïve Bayes Classifier for Large Scale Text," 2nd *International Conference on Artificial Intelligence: Technologies and Applications (ICAITA)*, Advances in Intelligent Systems Research, vol. 146, 2018.

[25] Bryn Farnsworth, "Facial Action Coding System (FACS) - A Visual Guidebook," August 18th, 2019.

[26] Hardianto Wibowo, Fachrunnisa Firdaus, Wildan Suharso, Wahyu Andhyka Kusuma, and Dani Harmanto, "Facial expression recognition of 3D image using facial action coding system (FACS)," Telkomnika, vol. 17, no. 2, pp628-636, 2019.

[27] Fabrizio Lamberti, Valentina Gatteschi, Andrea Sanna and Alberto Cannavo, "Animation Based on Live Performance and Natural Language Processing," International Journal of Human-Computer Interaction, vol. 35, Issue 18, 2019.

[28] Aditya Jain, Gandhar Kulkarni, and Vraj Shah, "Natural Language Processing," International Journal of Computer Science and Engineering (IJSE), vol. 6, Issue 1, pp161-167, 2018.

[29] Arifin, Surya Sumpeno, Muljono, Mochamad Hariadi, "A Model of Indonesian Dynamic Visemes From Facial Motion Capture Database Using A Clustering-Based Approach," IAENG International Journal of Computer Science, vol. 44, no. 1, pp41-51, 2017.