Asia-Pacific Journal of Information Technology and Multimedia 13(1) 2024: 01 - 20
https://doi.org/10.17576/apjitm-2024-1301-01

Recognizing Facial Emotion in Real-Time Using MuWNet a Novel Deep Learning Network

Pengecaman Emosi Wajah dalam Masa Nyata Menggunakan Kaedah Rangkaian Pembelajaran Mendalam MuWNet

Mustafa Mohammed Kataa, Wandeep Kaur*

Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor

**Corresponding author: mustafarh0@gmail.com*

Received 5 July 2023
Accepted 6 November 2023, Available online 1 June 2024

ABSTRACT

Facial expression recognition (FER) is a branch of psychology that studies the classification of human emotions using facial expressions. Particularly, FER can be implemented in a vast array of applications, including education, online entertainment, and even essential fields involving human lives and behavior, such as medicine. There are seven universal facial expression categories: surprise, sadness, happiness, contempt, fear, anger, and neutrality. Recognizing all these facial expressions and predicting a person's present mood is a challenging problem for machines. Because of the nature of humans, this challenge presents itself to a computer in a more sophisticated manner. The main objective of this research was to construct a novel deep Convolutional Neural Network (CNN) for facial expression classification that can assist in extracting features from images to identify facial gestures and then apply it in real-time. Various neural network models and classification methods have been introduced in the past to reach cutting-edge accuracy in this industry. Separate studies have investigated the capabilities and effectiveness of CNN models in distinguishing human emotions on the FER2013 dataset. In this study, the proposed MuWNet model has been diversified with several types of layers, such as convolution layers, separable convolution layers, and residual blocks. In addition, applying hyperparameter tweaking to enhance progress. The results of two experiments that have been done on the MuWNet model indicate that the accuracy of the classification in the second experiment was 70.72%, with an increase of 0.14% over the first. Finally, these results appear to be competitive in the field of FER, and it can be stated that the proposed model contributed to the emergence of a classification system for facial expressions.

Keywords**:** Facial Emotion recognition (FER), Convolutional neural network, Deep learning, FER2013 Dataset, Real-Time.

ABSTRAK

Pengecaman ekspresi muka (FER) adalah salah satu cabang psikologi yang mengkaji klasifikasi emosi manusia menggunakan ekspresi muka. FER boleh dilaksanakan dalam pelbagai bidang seperti Pendidikan, hiburan dalam talian, dan bidang penting yang melibatkan tingkah laku serta kehidupan manusia seperti perubatan. Secara universal terdapat tujuh kategori ekspresi muka iaitu terkejut, sedih, gembira, menghina, takut, marah, dan berkecuali. Mengenal pasti semua ekspresi muka ini dan meramalkan mood seseorang adalah masalah yang mencabar untuk mesin. Oleh kerana sifat semula jadi manusia yang pelbagai, cabaran ini muncul pada komputer dengan cara yang lebih kompleks. Objektif utama tesis ini adalah untuk membina Rangkaian Neural Convolutional (CNN) secara mendalam untuk mengklasifikasikan ekspresi muka yang boleh membantu dalam mengekstrak ciri daripada imej untuk mengenal pasti gerak isyarat muka dan kemudian menerapkannya dalam masa nyata. Terdapat pelbagai model rangkaian saraf dan kaedah klasifikasi telah diperkenalkan pada masa lalu untuk mencapai ketepatan termaju dalam industri ini. Kajian berasingan telah dilakukan untuk menyiasat keupayaan dan keberkesanan model CNN dalam membezakan emosi manusia pada set data FER2013. Dalam kajian ini, model MuWNet yang dicadangkan telah dipelbagaikan dengan beberapa jenis lapisan, seperti lapisan lilitan yang boleh dipisah menggunakan residual blocks. Di samping itu, tweaking hyperparameter turut digunakan untuk meningkatkan tahap kemajuan proses. Kajian ke atas dua eksperimen yang telah dilakukan pada model MuWNet jelas menunjukkan ketepatan pengelasan dalam eksperimen kedua iaitu 70.72%, dengan peningkatan sebanyak 0.14% berbanding dengan eksperimen pertama. Secara keseluruhannya, hasil dapatan kajian jelas menunjukkan bahawa penggunaan MUWNET adalah berdaya saing dalam bidang FER dan mampu menyumbang kepada sistem klasifikasi untuk mengenal pasti ekspresi muka.

Kata kunci: Pengecaman Emosi Muka (FER), Rangkaian saraf Konvolusi, Pembelajaran mendalam, Set Data FER2013, Masa Nyata.

INTRODUCTION

Emotions are a person's means of expressing their sentiments since people can communicate with one another either verbally or non-verbally. Hence, due to advances in technology, computer hardware, and graphics processing units (GPUs), the demand for human-computer interaction (HCI) has grown in recent decades. Researchers are now able to establish or construct a powerful artificial intelligence (AI) system that can automate a person's actions in a variety of industries.

The face expresses an individual's identity. Categorizing emotions through facial expressions can be more accurate than speaking or gesturing. Joy, neutrality, anger, fear, sadness, disgust, and surprise are seven facial expressions used by the authors (Pathar et al. 2019) to classify people's emotions. Similarly, in some cases, the interaction of the mouth, cheeks, eyes, brows, and front face could reveal more information about human feelings than words (Kaviya & Arumugaprakash 2020). Herein lies the importance of using FER in commerce, health, and education. Even though building a model to detect emotions is challenging, (Khaireddin & Chen 2021) stated that applying convolutional neural networks (CNNs) to this task could surpass other models that use classical image processing methods owing to the ability of CNNs to extract features from images and the effectiveness of their computation. Furthermore, when compared to traditional machine learning (ML) models such as support vector machines (SVM), CNN can produce high accuracy results (Gaddam et al. 2022).

In this research, a face recognizer will be implemented using deep learning neural networks. Finally, a real-time face recognition system will be used to identify human expressions in two phases. The first phase is face localization, which involves finding a face in an image or video. The second stage involves categorizing facial expressions into one of seven groups.

Knowing that building a human face expression recognizer using deep learning (DL) neural networks is a difficult task, as there are many important factors to consider, such as storage size, number of parameters, and layer level in a DL model, all of which can affect performance in real-time applications. When using several layers, as in AlexNet and VGGNet, which have a very deep structure, the complexity and size of CNNs grow, posing issues for real-time systems (Cotter 2020). This study tries to address the issue of facial emotion identification by developing a novel CNN model called MuWNet. While attempting to get a comparable accuracy outcome to state-of-the-art models.

LITERATURE REVIEW

This current study focuses on FER, which is the process of determining which expression is employed in a captured image or video. The review emphasizes the importance of comprehending human emotions. The vast majority of FER approaches incorporate the utilization of CNN and machine learning. Many different algorithms have also been used for a broad variety of datasets, such as FER-2013, the CK+, the RaFD, the JAFFE datasets, and many others.

Sang et al. (2017) developed multiple methods for identifying facial expressions in humans. The methods are dependent on CNN. Their techniques are influenced by the VGG design principles. In facial expression recognition, the authors found that L2 multi-class SVM loss is preferable to cross-entropy loss by comparing them on the FER2013 dataset using different loss functions. BKVGG12, which consisted of 12 layers, was the model with the highest level of accuracy, coming in at 71.9%.

Using the FER-2013 dataset, Agrawal and Mittal (2019) built two CNN algorithms. More, they evaluated the influence of CNN parameters, notably kernel size and filters' number, on the classification precision. Their work made a substantial contribution by testing a variety of kernel sizes along with filters to propose two new CNN architectures with a 65% human-like accuracy. According to their research, the kernel size and filters' number were found to have a substantial effect on the network's accuracy. In addition, the accuracy of both proposed models exceeded 65%.

Bhandari and Pal (2021) investigated whether the use of edges can help CNN identify emotions from images. To identify facial expressions from photographs, a CNN model consisting of two towers and accepting a variety of inputs has been developed. Accordingly, they reasoned that edges in an image provide discriminatory information and that their explicit usage is predicted to assist in the training of CNNs and better emotion recognition. This is because their explicit use is expected to help improve accuracy. Their proposed CNN included two inputs. The first input was the image itself, while the second was the edge image acquired by the Canny edge detector. The researchers tested their findings on two datasets, JAFEE and FER2013, to show that using edge information explicitly enhances classifier performance. The proposed model attained an accuracy of 85.7% on the JAFEE dataset and 63.7% on the FER2013 dataset.

Vulpe-Grigorasi and Girgore (2021) aimed to increase the efficiency of a CNN network by fine-tuning its architecture and hyperparameters to classify human facial images into distinct emotional categories. They did this by using images of people's faces. The ideal hyperparameters were identified by creating and training models utilizing a random search technique applied to a search space containing discrete hyperparameter values. To prove that an effective solution may be discovered in a search space where previous results are considered to be local minima, the researchers wanted to optimize the hyperparameters of the model superficially. In addition, the only layers of the model architecture that were optimized were the convolutive layers; the categorization layers were left out of the optimization process. The most accurate model was able to attain a score of 72.16 % after being trained and evaluated using the FER2013 database.

As several applications respond to the emotional state of a participant, it is crucial to establish an accurate FER on a smartphone. Cotter (2020) introduced MobiExpressNet, a novel light deep-learning model for FER based on two frameworks, MobileNetV1 and MobileNetV2. The researcher began by utilizing a series of kernels and then moved on to utilizing a series of depth-wise convolution filters, to extract the feature maps. The investigation led to the discovery by the researcher that the best network model has an accuracy of 67.96% of the challenging FER2013 dataset. This is 2.5% more accurate than human accuracy. In addition, it was discovered that the MobiExpressNet model was more than five times smaller than the smallest MobileNet model, which makes the new model particularly interesting for use in real-time applications.

RESEARCH METHODOLOGY

The methodology utilized in this study is broken down into four distinct phases, the first of which entails the preparation of the FER2013 dataset along with its preliminary processing. The second phase is known as the training phase, where the suggested model is trained, and modifications to the ideal parameters are determined. In the third phase, the created model is evaluated using the currently available data. To assess its performance and acquire the required level of precision, it is necessary to compare the proposed model to the research that came before it. The final phase of this research involves putting the model into action in real-time to determine whether it can assist in the identification of human emotions. Figure 1 provides a graphic representation of the study design.

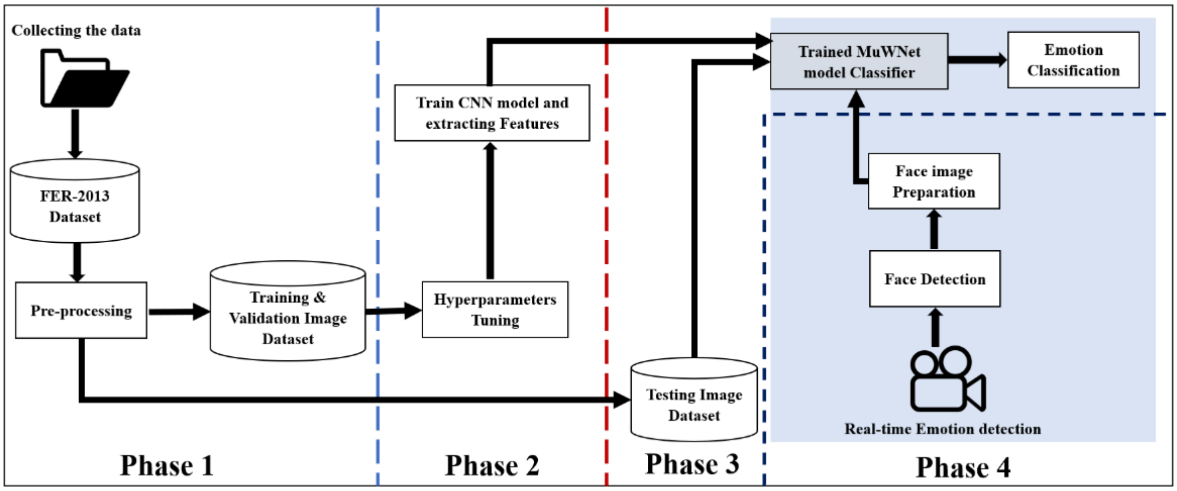


FIGURE 1. Graphical Representation of The Proposed System

As previously indicated, this research is separated into four phases. The first phase consists of two steps. The first involves collecting and preparing the data, while the second entails pre-processing the data and separating it into training, validation, and testing datasets.

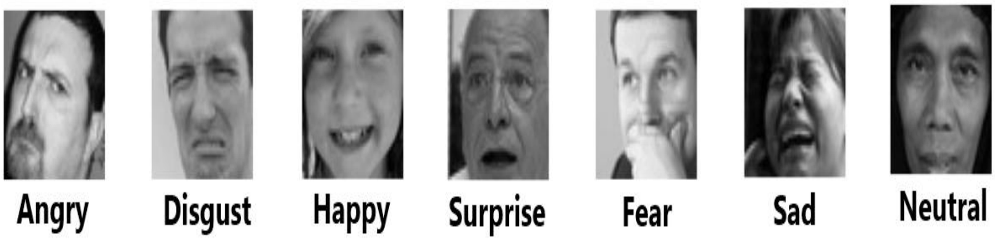
The second phase includes three steps the initial step entails collecting the training data. The subsequent step is hyperparameter tuning, which involves locating the optimal hyperparameters to enhance model accuracy on image data. In the final step, CNN, one of the DL methods, will be utilized to train the model using an image training dataset. CNN will assist in the extraction of characteristics to be included in the proposed model, removing the need to manually extract them.

The third phase employs the testing data images to gauge the model's performance and compare its accuracy to that of similar works. This phase is comprised of two distinct parts. Having available test data to feed into the proposed model is the initial step, followed by assessing the mentored model on test images and determining its prediction accuracy.

In the fourth phase, the MuWNet model will be applied in real-time to assist in recognizing human emotions. This might be accomplished by first employing the Haar Cascade classifier, a well-known face recognition technique, which will be utilized in this study to capture a human face from a video frame and crop the face so that it can be fed on the suggested model. Secondly, a face image preparation will be conducted to resize the cropped face image so that it can be given as input to the proposed model. Finally, the MuWNet model, which was trained and tested in phases 2 and 3, is now ready to be provided with the resized image, and the output is the category to which the emotion belongs.

1. Dataset

The FER2013 dataset used in this study was created by Carrier and Courville (2013) as part of a larger research project and was used in one of Kaggle's representation learning competitions. The FER2013 data set includes grayscale facial images with 48 by 48 pixels and seven numerically ordered classes, where Angry is represented by 0, Disgust by 1, Fear by 2, Happy by 3, Sad by 4, Surprise by 5, and Neutral by 6. The training set comprises 28,709 distinct examples for each of the seven categories. For the leaderboard, a public test set consisting of 3,589 samples was employed. Furthermore, to identify the contest's winner, an extra 3,589 samples were incorporated into the final private test set. The FER-2013 dataset is represented in its entirety by Figure 2 along with some samples.



Source: Carrier & Courville 2013

FIGURE 2. Samples From the FER2013 Dataset

2. MUWNET Model Architecture

This experiment aims to develop the MuWNet model so that it can recognize seven distinct facial emotions.

The deep neural network that is a component of the MuWNet model drew its motivation from that of the VGGNet network (Simonyan & Zisserman 2014), the ResNet network (He et al. 2016), and the MobileNet (Howard et al. 2017). Also, the MuWNet model was named after the student and supervisor for this study, with Mu being the first two letters of Mustafa and W being the first letter of Wandeep. The structure of the MuWNet CNN model is depicted in Figure 3.

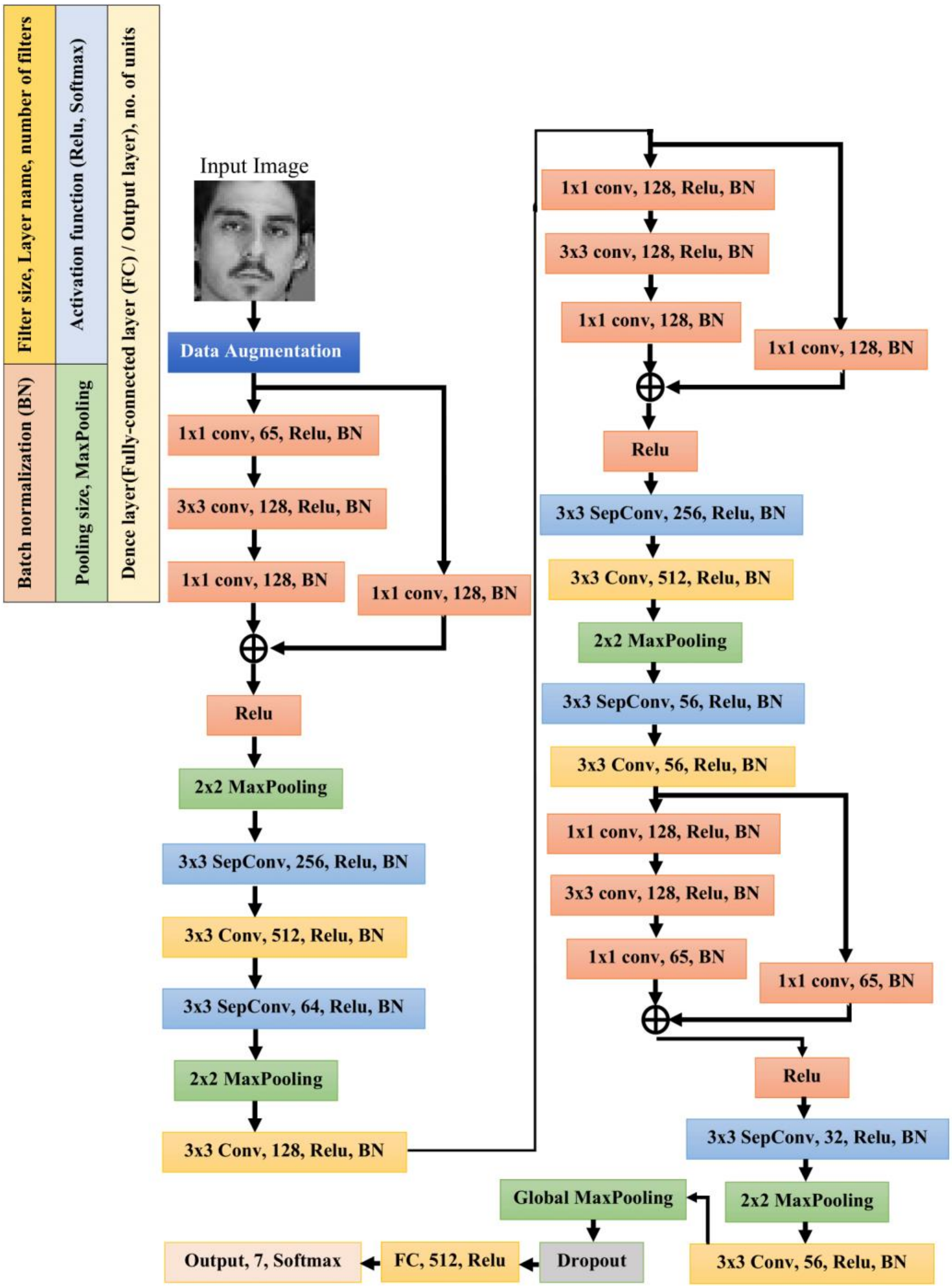


FIGURE 3. The MUWNET CNN Model Architecture

Data augmentation pre-processing was performed within the model to make these techniques active only during training. These approaches include horizontal flip, adjusting the width and height by ±20°; zooming by ±15°; rotation by ±15°; and finally, normalizing the image pixel values by dividing them by 255. Figure 4 displays several examples of data augmentation on a random image. Plus, a batch normalization (BN) layer has been placed after each convolution layer or the separable convolution layer as it helps stabilize training and lower the ultimate loss which makes the loss curve in neural networks much more stable.

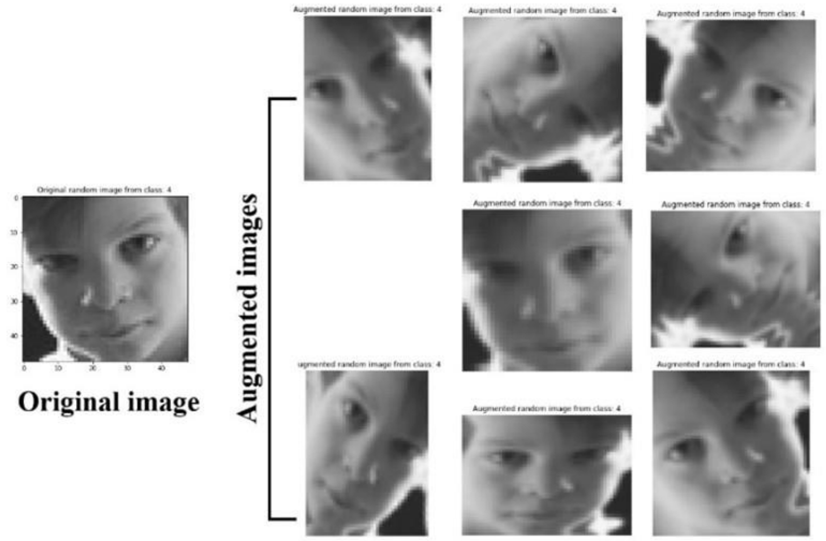

Augmented random image from class: 4
Augmented random image from class: 4
Augmented random image from class: 4
Original random image from class: 4
0
10
20
30
40
0
10
20
30
40
Original image
Augmented images
Augmented random image from class: 4
Augmented random image from class: 4
ugmented random image from class: 4
Augmented random image from class: 4
Augmented random image from class: 4

FIGURE 4. Examples of Augmented Image

3. Evaluation

Therefore, the evaluation procedure demonstrates how well the model applied in this study performed. Along with the accuracy/loss plot, the study used performance indicators such as accuracy, recall, precision, and f1-score. Additionally, plotted accuracy versus loss.

ENVIRONMENT EXPERIMENTAL SETTINGS

At the beginning of this research project, the data processing was done with Python. In addition, this investigation made use of Google CoLab as the foundational environment because it enables us to access a robust graphics processing unit (GPU) and makes it straightforward for us to construct and validate the proposed model. Specifications for Google CoLab: GPU Tesla P100-PCIE-16 GB, Random Access Memory (RAM) 12 GB, and Hard Disk Drive (HHD) 124 GB.

RESULTS & DISCUSSION

HYPERPARAMETERS SELECTION

Based on previous research and knowledge of hyperparameters, this research employed a specific set of hyperparameters and established a range of values for examination. In (Vulpe-Grigorasi and Grigore 2021), the authors applied the dropout hyperparameter to demonstrate how different dropout values could alter model performance. Additionally, (Agrawal and Mittal 2019) utilized a variety of optimizers to check their influence on model accuracy compared to the ADAM optimizer, which is considered a default optimizer. Moreover, the remaining hyperparameters—image size, batch-size, and fully connected layer—were chosen for this study to investigate their impact on model performance. Finally, distinct values within a defined range were produced for each parameter. Table 1 presents a summary of the value ranges employed in each experiment supported by the number of epochs.

TABLE 1. Values of The Hyperparameters

Hyperparameter	Value Range	No. of epochs
Image Size	[48, 64, 128, 150]	25
Optimizers	['Adam', 'SGD', 'RMSprop']	25
Batch_Size	[32, 64, 128]	20
Fully connected layer	[32, 64, 128, 512, 1028]	25
Dropout	[0.1, 0.2, 0.3, 0.4, 0.5]	20

After uploading the FER2013 dataset to Google Drive, the experiments ran on Google CoLab. Each best hyperparameter was determined independently see Appendix I, which is done by running an experiment for each hyperparameter, beginning with image size, and ending with dropout.

The results were acquired by tweaking the hyperparameters of the proposed CNN-based model classifier. The evaluation was performed by summing the validation accuracy for each epoch and dividing the entire sum by the epochs' number specified for each parameter. The average validation for both accuracy and loss for each parameter using the MuWNet model network are shown in Table 2. Figure 5 demonstrates the accuracy/loss plot of the hyperparameters for each epoch.

TABLE 2. Values of The Hyperparameters

Hyperparameter	No. of epochs	Values	Average validation accuracy	Average validation loss
Image Size	25	48×48	42.70%	147.15
		64×64	44.03%	143.44
		128×128	50.86%	127.77
		150×150	51.92%	125.05
Optimizers	25	'SGD'	42.69%	147.99
		'RMSprop'	48.26%	139.41
		'Adam'	51.92%	126.46
Batch_Size	20	32	50.15%	131.23
		64	49.66%	132.11
		128	47.66%	137.07
Fully connected layer	25	32	51.16%	126.56
		64	51.26%	126.97
		128	52.23%	124.87
		512	53.49%	121.25
		1028	53.33%	121.78
Dropout	20	0.1	48.34%	136.28
		0.2	49.46%	133.34
		0.3	48.36%	135.60
		0.4	46.76%	138.18
		0.5	45.96%	139.77

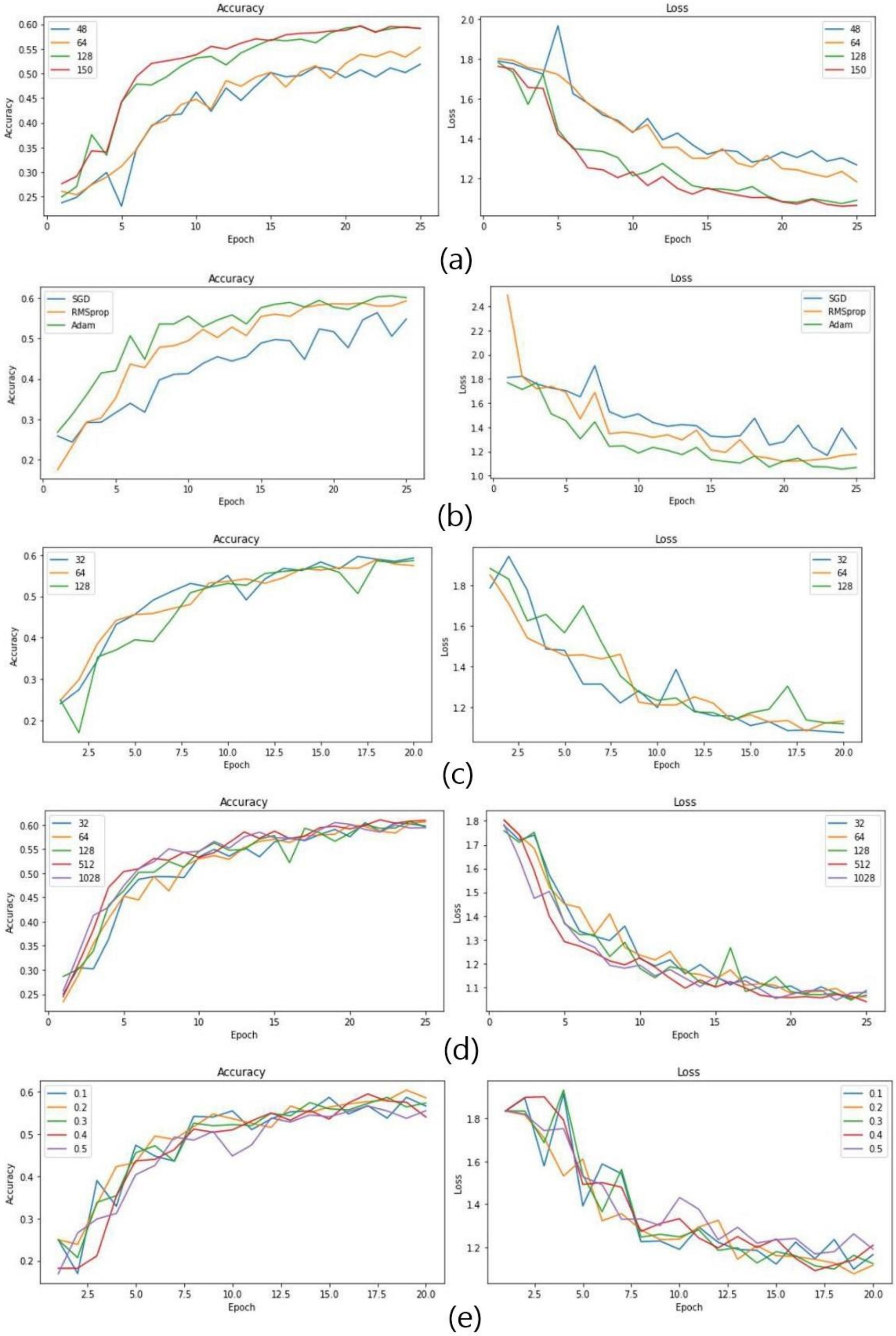


FIGURE 5. The accuracy/loss plot for each hyperparameter where: (a) Image Size, (b) Optimizers, (c)Batch_Size, (d) Fully connected layer, and (e) Dropout

Table 3 shows the model's final parameters after the tuning trials for the hyperparameters were done.

TABLE 3. The Selected Parameters For The MUWNET Model

Image Size	Optimizers	Batch_Size	Fully connected layer	Dropout
150×150	'Adam'	32	512	0.2

Furthermore, in this study, the learning rate was adjusted using the ReduceLROnPlateau offered by Keras callbacks. This callback aims to fine-tune model weights by slowing down the rate of learning when the model performance stops improving.

MUWNET MODEL RESULTS ON THE FER2013 DATASET

The proposed model was used to conduct two experiments in order to acquire optimal outcomes. In the first experiment, the MuWNet model was trained using training data. In addition, the learning rate began at 0.001 and was changed throughout training using ReduceLROnPlateau. This strategy was used to fine-tune model weights and overcome the issue of overfitting. The model was initially trained for 150 epochs. ReduceLROnPlateau decreased the learning rate from 0.001 to 0.0001 at epoch 135 during training, as shown in Figure 6. After 150 epochs, the validation accuracy (PublicTest) was 68.10% and the test data accuracy (PrivateTest) was 70.41%.

To improve the findings, the MuWNet model was trained for an additional 50 iterations. This demonstrated a shift in the learning rate from 0.0001 to 0.00001 at epoch 187, and at the end of these epochs, the model enhanced the outcomes by 68.43% for the validation data and by 70.52% for the test data.

In addition, looking at Figure 7, the model showed overfitting behavior with a training loss of 0.5612 and a validation loss of 0.9696 at epoch 200. As a result, 50 additional epochs were executed to surpass the training results and attempt to minimize the loss values for both training and validation to achieve the optimal fit. Finally, the results were improved after training the model for 250 epochs, with the model's accuracy on test data being 70.58% and a learning rate of 0.00001.

The accuracy/loss plots for this experiment are illustrated in Figures 6 and 7. The accuracy/loss plots showed that the learning rate changed by 0.0001 and 0.00001 at epochs 135 and 187, respectively. After epoch 187, the model's accuracy and loss almost stopped changing because it could not eliminate the problem of overfitting. That made the model too complicated to learn more from the same set of data. Table 4 presents the performance metrics for the first experiment.

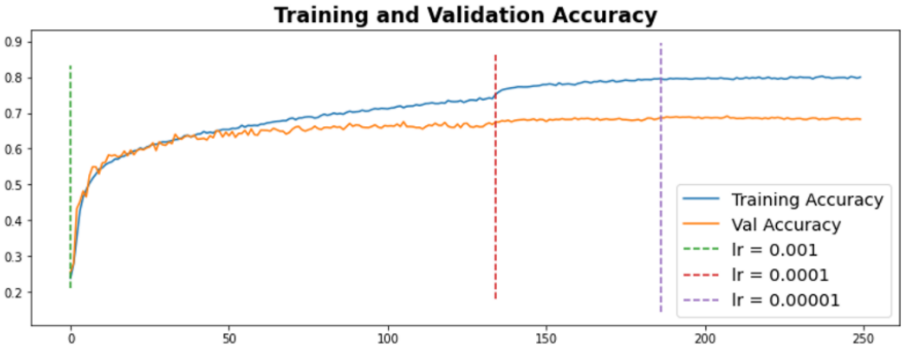


FIGURE 6. First Experiment Accuracy Plot for MUWNET Model on FER2013 Dataset

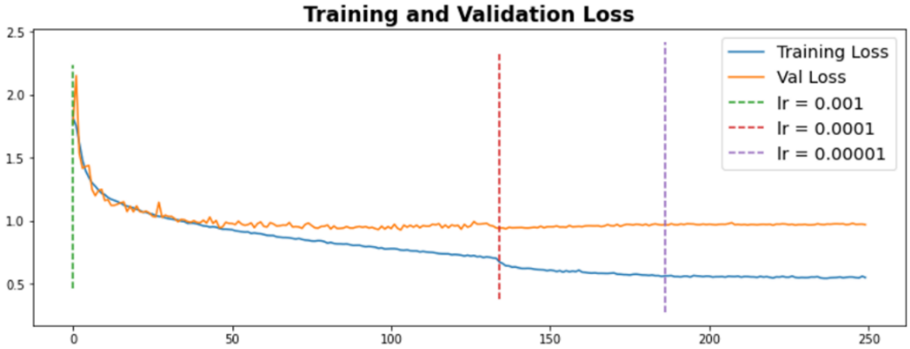
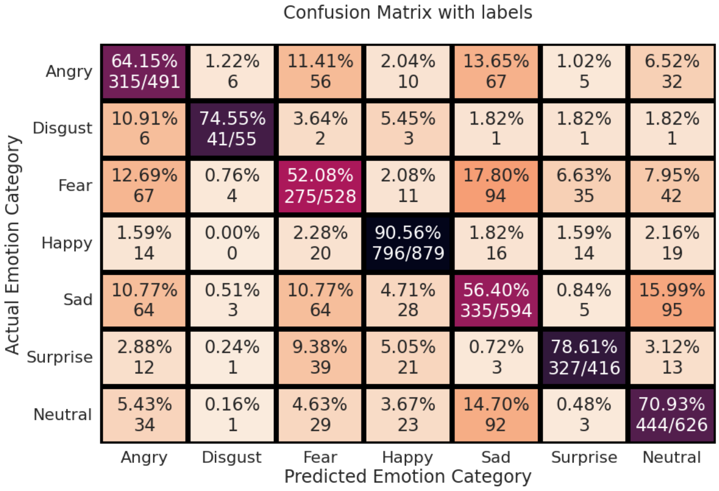


FIGURE 7. First Experiment Loss Plot for MUWNET Model on FER2013 Dataset

TABLE 4. Performance Metrics for The First Experiment

Accuracy	Recall	Precision	F1-Score
70.58%	69.61%	69.76%	69.65%

To comprehend how the model in this experiment categorized the seven distinct emotions, the confusion matrix was employed in Figure 8.


Confusion Matrix with labels

Actual Emotion Category	Angry	Disgust	Fear	Happy	Sad	Surprise	Neutral
Angry	64.15% 315/491	1.22% 6	11.41% 56	2.04% 10	13.65% 67	1.02% 5	6.52% 32
Disgust	10.91% 6	74.55% 41/55	3.64% 2	5.45% 3	1.82% 1	1.82% 1	1.82% 1
Fear	12.69% 67	0.76% 4	52.08% 275/528	2.08% 11	17.80% 94	6.63% 35	7.95% 42
Happy	1.59% 14	0.00% 0	2.28% 20	90.56% 796/879	1.82% 16	1.59% 14	2.16% 19
Sad	10.77% 64	0.51% 3	10.77% 64	4.71% 28	56.40% 335/594	0.84% 5	15.99% 95
Surprise	2.88% 12	0.24% 1	9.38% 39	5.05% 21	0.72% 3	78.61% 327/416	3.12% 13
Neutral	5.43% 34	0.16% 1	4.63% 29	3.67% 23	14.70% 92	0.48% 3	70.93% 444/626

FIGURE 8. Confusion Matrix for MUWNET Model in The First Experiment on The FER2013 Dataset

In the second experiment, training and validation data were combined into a single training dataset. The MuWNet model was then trained for 150 epochs. The model's accuracy on validation data was 79.86%, and on test data it was 68.01%.

The accuracy of the test data was less than what was achieved by the same model in the first experiment, and the learning rate's value did not alter during training; therefore, the value of the learning rate was modified manually from 0.001 to 0.0001 to train the model for an extra 50 epochs. After completing the additional 50 epochs, the model obtained an accuracy of 85.71% on validation data and 70.72% on test data, respectively. In the second experiment, the MuWNet model achieved loss values of 0.5412 for training and 0.4110 for validation at epoch 200, as shown in Figure 9 and 10.

Fifty training epochs were added to the model's total epochs to enhance the performance and reduce the training and validation loss values. By the time the training was over, the model had barely changed, and the test data accuracy had stayed at 70.72%. At the same time, the training and validation loss values were reduced to 0.5076 and 0.3738, respectively, which indicates that the model is learning and requires extra training epochs to alter its accuracy value. Figures 9 and 10 reveal this experiment's accuracy and loss plots. Furthermore, the confusion matrix in Figure 11 was used to understand how the model in this experiment classified the seven various emotions, and Table 5 presents the performance metrics for the second experiment.

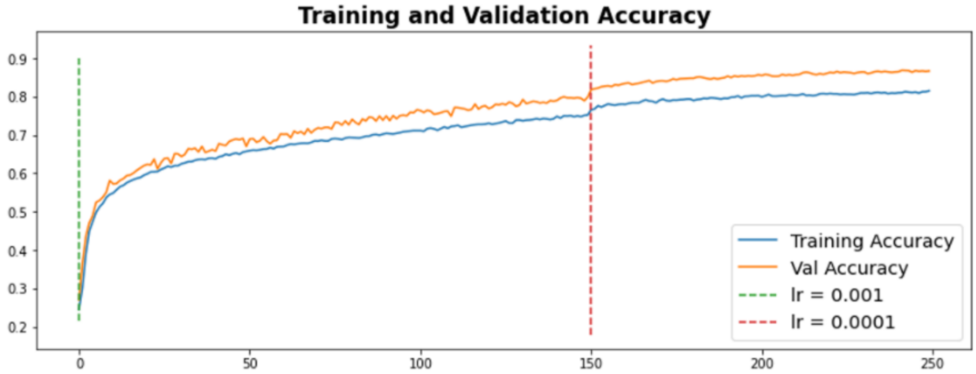


FIGURE 9. Second Experiment Accuracy Plot for MUWNET Model on FER2013 Dataset

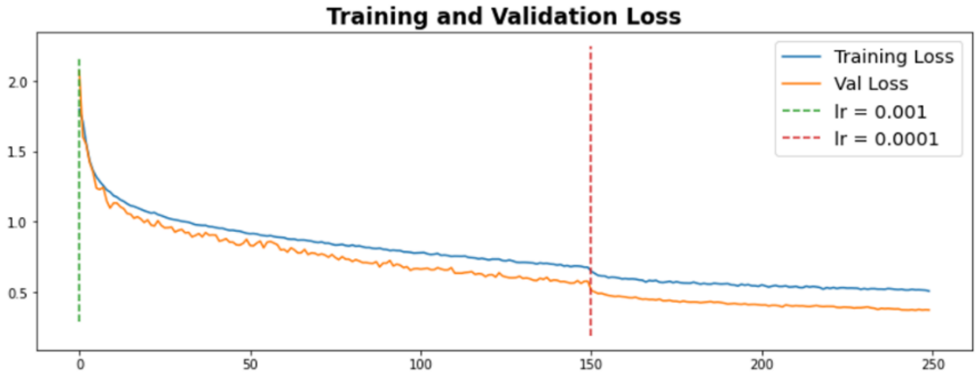
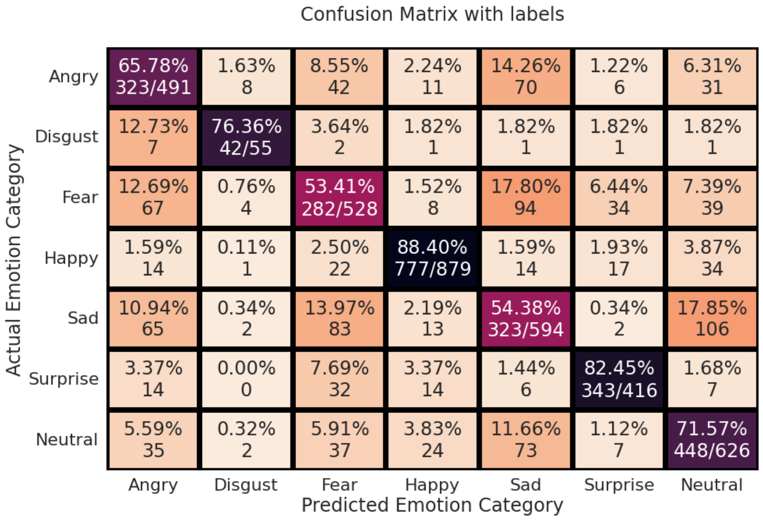


FIGURE 10. Second Experiment Loss Plot for MUWNET Model on FER2013 Dataset

TABLE 5. Performance Metrics for The Second Experiment

Accuracy	Recall	Precision	F1-Score
70.72%	70.34%	69.61%	69.93%



Confusion Matrix with labels

Actual Emotion Category	Angry	Disgust	Fear	Happy	Sad	Surprise	Neutral
Angry	65.78% 323/491	1.63% 8	8.55% 42	2.24% 11	14.26% 70	1.22% 6	6.31% 31
Disgust	12.73% 7	76.36% 42/55	3.64% 2	1.82% 1	1.82% 1	1.82% 1	1.82% 1
Fear	12.69% 67	0.76% 4	53.41% 282/528	1.52% 8	17.80% 94	6.44% 34	7.39% 39
Happy	1.59% 14	0.11% 1	2.50% 22	88.40% 777/879	1.59% 14	1.93% 17	3.87% 34
Sad	10.94% 65	0.34% 2	13.97% 83	2.19% 13	54.38% 323/594	0.34% 2	17.85% 106
Surprise	3.37% 14	0.00% 0	7.69% 32	3.37% 14	1.44% 6	82.45% 343/416	1.68% 7
Neutral	5.59% 35	0.32% 2	5.91% 37	3.83% 24	11.66% 73	1.12% 7	71.57% 448/626

FIGURE 11. Confusion Matrix for MUWNET Model in The Second Experiment on The FER2013 Dataset

In the second experiment, the combined dataset from the training and validation datasets exposed the model to more instances. In addition, feeding this dataset to the model reduced the gap between training and validation accuracy and loss values during training, which helped the MuWNet model perform significantly better at classifying emotions than it did in the first experiment, with an increase of 0.14% in the accuracy value. Figure 12 demonstrates the variance in the classification of emotions between the two MuWNet model experiments.

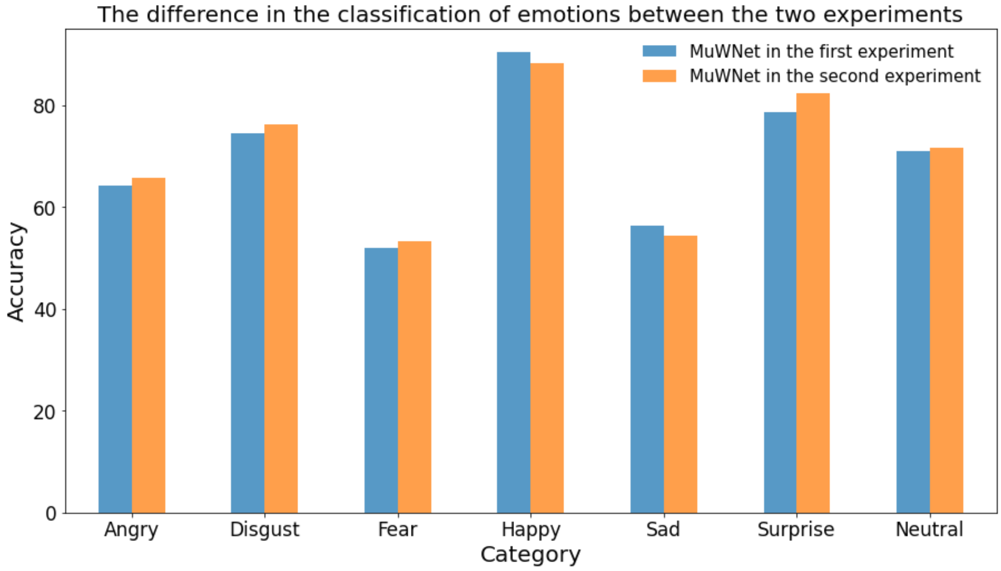


FIGURE 12. The Difference in The Classification of Emotions Between the Two Experiments

Table 6 shows how the model from the first experiment was compared to models from other studies that also used the FER2013 dataset.

TABLE 6. The Comparison Between The Proposed Model And Other Studies

Author(s)	Accuracy on FER2013 Dataset	No. of parameters
Yanling Gan et al. 2019	73.73%	—
Khaireddin & Chen, 2021	73.28%	—
Abbassi et al. 2020	72.806%	—
Vulpe-Grigorasi & Grigore 2021	72.16%	5.17M
Sang et al. 2017	71.9%	4.19M
Shao & Qian 2019	71.14%	7.12M
The Proposed MUWNET Model	**70.58%**	**3.1M**
Onyema et al. 2021	70%	—
Cotter 2020	67.96%	75,079
Agrawal & Mittal 2019	65.77%	0.93M
Yijun Gan 2018	64.24%	—
Bhandari & Pal, 2021	63.7%	9.9M
Gaddam et al. 2022	55.6%	—

By comparing the result to the prior research presented in Table 6, it can be declared that the model can reasonably recognize distinct facial expressions. In the same vein, the MuWNet utilizes a single input, which is the image itself, contrary to the work of Bhandari and Pal (2021), which included edges as an extra input. The reason for using a single input is that the model converts the original image into edges during the process, so providing it with edges would not be advantageous.

Even though the MuWNet model did not produce a breakthrough result in comparison to previous studies, the suggested model was trained without any prior information, unlike the work of Yijun Gan (2018) and Gaddam et al. (2022), in which transfer learning was used relying on AlexNet and ResNet50 respectively. Adding to that, the accuracy acquired from the MuWNet model was superior to both studies.

Furthermore, the MuWNet model adopted many layer types and architectures that strengthened and diversified the model structure, such as convolution layers, separable convolution layers, and residual blocks, in contrast to (Abbassi et al. 2020, Agrawal & Mittal 2019, Bhandari & Pal 2021), who employed only one type of layer in their investigations. Having these layers in the proposed model could empower it, as convolution layers allow the model to automatically identify meaningful characteristics, adding separable convolution layers may help in reducing network parameters without compromising accuracy, and the use of residual blocks may help handle the issue of vanishing and bursting gradients.

Moreover, the suggested model has 3.1 million parameters, which is fewer than each of the other models' (Sang et al. 2017; Shao & Qian 2019; Vulpe-Grigorasi & Grigore 2021) parameters, making it more suitable to operate in real-time since more parameters demand more computations, which take longer time. At this current model performance, the MuWNet model resulting from the second experiment was applied in real-time, and some real-time examples are presented in Figure 13.

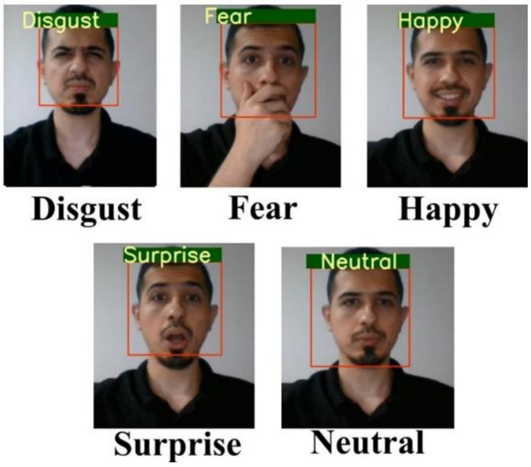


FIGURE 13. Samples of Real-Time Using MUWNET Model

MUWNET MODEL STATISTICAL ASSESSMENT

In order to comprehensively assess the MuWNet model, the standard deviation, z-score, and mean were calculated using the data from Table 6. Below are the formulae that were used to determine each of the aforementioned variables.

$$\mu = \frac{\sum \chi}{N} \tag{1}$$

Where:

μ	*: average accuracy (mean).*
χ	*: accuracy value.*
$\sum x$	*: sum of each x value.*
N	*: number of summed accuracy values.*

(Source: Jason 2020)

$$\sigma = \sqrt{\frac{\sum(\chi - \mu)^2}{N}} \tag{2}$$

Where:

σ	*: standard deviation.*
χ	*: accuracy value.*
μ	*: mean value.*
N	*: number of accuracy values.*

(Source: Joydeep 2017)

Utilizing the previously mentioned formulas, the resulting mean and standard deviation values were 68.682 and 4.978, respectively. After determining these variables, the z-score for each model accuracy was computed using the equation below and results are shown in Table 7.

$$z = \frac{\chi - \mu}{\sigma} \tag{3}$$

Where:

z *: z-score.*
σ *: standard deviation.*
χ *: accuracy value.*
μ *: mean value.*

(Source: Mcleod 2023)

TABLE 7. z-score value for each model's accuracy

Author(s)	Accuracy on FER2013 Dataset	z-score
Yanling Gan et al. 2019	73.73%	1.013
Khaireddin & Chen, 2021	73.28%	0.923
Abbassi et al. 2020	72.806%	0.829
Vulpe-Grigorasi & Grigore 2021	72.16%	0.698
Sang et al. 2017	71.9%	0.646
Shao & Qian 2019	71.14%	0.493
The Proposed MUWNET Model	70.58%	0.381
Onyema et al. 2021	70%	0.264
Cotter 2020	67.96%	-0.145
Agrawal & Mittal 2019	65.77%	-0.584
Yijun Gan 2018	64.24%	-0.892
Bhandari & Pal, 2021	63.7%	-1.000
Gaddam et al. 2022	55.6%	-2.627

Within the provided table, the z-score serves as a statistical measure that quantifies the deviation of a given accuracy score from the average accuracy score. A positive z-score indicates that the accuracy score surpasses the mean, while a negative z-score implies that the accuracy score falls short of the mean.

In this instance, the proposed MuWNet model exhibits a z-score of 0.381, indicating that its accuracy score oversteps the mean accuracy score by 0.381 standard deviations. This implies that the proposed model exceeds the performance of the other models listed in the table. In simpler terms, the proposed model possesses a higher likelihood of correctly classifying facial expressions compared to models that fall below average accuracy.

CONCLUSION

This research aims to investigate the techniques and features used for facial emotion recognition in images using a novel DL model called MuWNet for classifying seven distinct facial emotions. In addition, the results obtained were compared with those from previous studies, confirming that the model can accurately organize emotions. The categorization of the FER2013 dataset served as the model's challenge for this study. Herein, the MuWNet model was evaluated in real-time to capture human emotions.

This study was aided by the adaptation of several contemporary DL techniques and the usage of several layer types, such as convolution layers, separable convolution layers, and residual blocks. Moreover, employing hyperparameter adjustment was a good technique to boost the performance of various models, according to the findings of related works. Different layers, such as the BN Layer and Dropout Layer, were utilized in this investigation due to their usefulness in stabilizing training outcomes and minimizing overfitting, as reported by other works.

Accordingly, the suggested model contributed to the evolution of a face-expression classification system based on a DNN model. In addition, applying several hyperparameters and analyzing their influence on the suggested model.

Two experiments were conducted on the MuWNet model, and the findings show that the classification accuracy of the second experiment was 70.72%, with an increase of 0.14% over the first.

To further this research, evaluating diverse datasets like AffectNet and the Extended Cohn-Kanade Dataset (CK+), and exploring hyperparameter optimization with a grid search can provoke valuable insights and refine the model's accuracy. Furthermore, Due to the disparity in performance between classes, employing various kernel_initializer classes may yield fruitful results. Similarly, using a different kernel_regularizer, such as L1 or L2, could improve the model's performance. Moreover, developing an interactive facial recognition system integrated with a simple game can provide a fun and engaging platform for real-time emoticon selection based on facial expressions. Finally, incorporating speech processing as an auxiliary input alongside the image could help in resolving the illumination problems.

ACKNOWLEDGEMENT

This research was supported by Fakulti Teknologi dan Sains Maklumat, Universiti Kebangsaan Malaysia (UKM).

REFERENCE

Abbassi, Nessrine, Rabie Helaly, Mohamed Ali Hajjaji, and Abdellatif Mtibaa. "A Deep Learning Facial Emotion Classification System: A VGGNET-19 Based Approach." *2020 20th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA),* December 20, 2020. https://doi.org/10.1109/sta50679.2020.9329355.

Agrawal, Abhinav, and Namita Mittal. "Using CNN for Facial Expression Recognition: A Study of the Effects of Kernel Size and Number of Filters on Accuracy." *The Visual*

Computer 36, no. 2 (January 23, 2019): 405–12. https://doi.org/10.1007/s00371-019-01630-9.

Bhandari, Arkaprabha, and Nikhil R. Pal. "Can Edges Help Convolution Neural Networks in Emotion Recognition?" *Neurocomputing* 433 (April 2021): 162–68. https://doi.org/10.1016/j.neucom.2020.12.092.

Carrier , Pierre-Luc, and Aaron Courville. "Challenges in Representation Learning: Facial Expression Recognition Challenge." Kaggle, 2013. https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/data.

Cotter, Shane F. "MobiExpressNet: A Deep Learning Network for Face Expression Recognition on Smart Phones." *2020 IEEE International Conference on Consumer Electronics (ICCE)*, January 2020. https://doi.org/10.1109/icce46568.2020.9042973.

Gaddam, Dharma Karan, Mohd Dilshad Ansari, Sandeep Vuppala, Vinit Kumar Gunjan, and Madan Mohan Sati. "Human Facial Emotion Detection Using Deep Learning." *Lecture Notes in Electrical Engineering*, January 1, 2022, 1417–27. https://doi.org/10.1007/978-981-16-3690-5_136.

Gan, Yanling, Jingying Chen, and Luhui Xu. "Facial Expression Recognition Boosted by Soft Label with a Diverse Ensemble." *Pattern Recognition Letters* 125 (July 2019): 105–12. https://doi.org/10.1016/j.patrec.2019.04.002.

Gan, Yijun. "Facial Expression Recognition Using Convolutional Neural Network." *Proceedings of the 2nd International Conference on Vision, Image and Signal Processing*, August 27, 2018. https://doi.org/10.1145/3271553.3271584.

He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep Residual Learning for Image Recognition." *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016. https://doi.org/10.1109/cvpr.2016.90.

Howard, Andrew G., Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto and Hartwig Adam. "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications." *ArXiv* abs/1704.04861 (2017): n. pag.

Khaireddin, Yousif and Zhuo Liang Chen. "Facial Emotion Recognition: State of the Art Performance on FER2013." *ArXiv* abs/2105.03588 (2021): n. pag.

Jason, Brownlee. "Arithmetic, Geometric, and Harmonic Means for Machine Learning." MachineLearningMastery.com, August 19, 2020. https://machinelearningmastery.com/arithmetic-geometric-and-harmonic-means-for-machine-learning/.

Joydeep, Bhattacharjee. "Basics of Statistics for Machine Learning Engineers II." Medium, October 11, 2017. https://medium.com/technology-nineleaps/basics-of-statistics-for-machine-learning-engineers-ii-d25c5a5dac67.

Mcleod, Saul. "Z-Score: Definition, Formula, Calculation & Interpretation." Simply Psychology, October 6, 2023. https://www.simplypsychology.org/z-score.html.

Onyema, Edeh Michael, Piyush Kumar Shukla, Surjeet Dalal, Mayuri Neeraj Mathur, Mohammed Zakariah, and Basant Tiwari. "Enhancement of Patient Facial Recognition through Deep Learning Algorithm: ConvNet." *Journal of Healthcare Engineering* 2021 (December 6, 2021): 1–8. https://doi.org/10.1155/2021/5196000.

P, Kaviya, and Arumugaprakash T. "Group Facial Emotion Analysis System Using Convolutional Neural Network." *2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184)*, June 2020. https://doi.org/10.1109/icoei48184.2020.9143037.

Pathar, Rohit, Abhishek Adivarekar, Arti Mishra, and Anushree Deshmukh. "Human Emotion Recognition Using Convolutional Neural Network in Real Time." *2019 1st*

International Conference on Innovations in Information and Communication Technology (ICIICT), April 2019. https://doi.org/10.1109/iciict1.2019.8741491.

Sang, Dinh Viet, Nguyen Van Dat, and Do Phan Thuan. “Facial Expression Recognition Using Deep Convolutional Neural Networks.” *2017 9th International Conference on Knowledge and Systems Engineering (KSE),* October 2017. https://doi.org/10.1109/kse.2017.8119447.

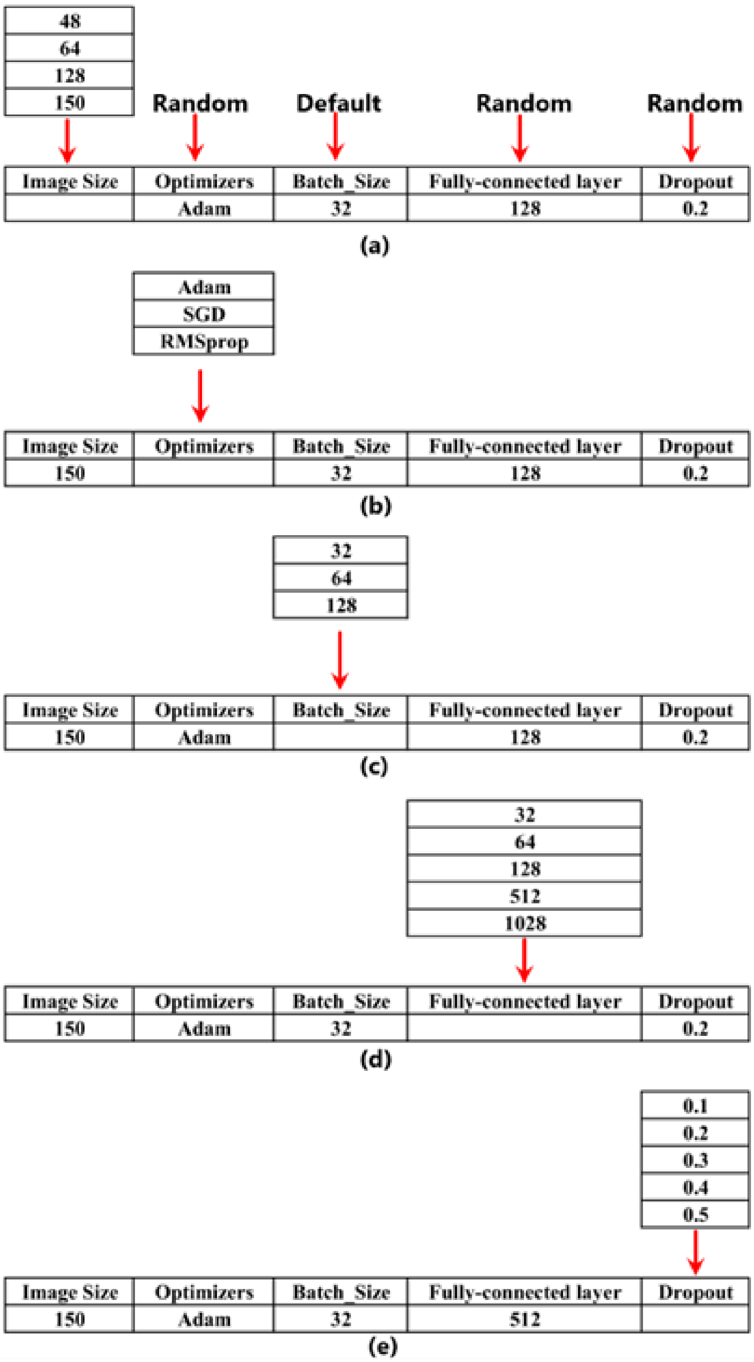
Shao, Jie, and Yongsheng Qian. “Three Convolutional Neural Network Models for Facial Expression Recognition in the Wild.” *Neurocomputing* 355 (August 2019): 82–92. https://doi.org/10.1016/j.neucom.2019.05.005.

Simonyan, K, and A Zisserman. 2015. “Very Deep Convolutional Networks for Large-Scale Image Recognition.” In 3rd International Conference on Learning Representations (ICLR 2015), 1–14. Computational and Biological Learning Society.

Vulpe-Grigorasi, Adrian, and Ovidiu Grigore. “Convolutional Neural Network Hyperparameters Optimization for Facial Emotion Recognition.” *2021 12th International Symposium on Advanced Topics in Electrical Engineering (ATEE),* March 25, 2021. https://doi.org/10.1109/atee52255.2021.9425073.

Appendix I

A Figure that shows how the hyperparameters were chosen.



48
64
128
150

Image Size	Optimizers	Batch_Size	Fully-connected layer	Dropout
	Adam	32	128	0.2

(a)

Adam
SGD
RMSprop

Image Size	Optimizers	Batch_Size	Fully-connected layer	Dropout
150		32	128	0.2

(b)

32
64
128

Image Size	Optimizers	Batch_Size	Fully-connected layer	Dropout
150	Adam		128	0.2

(c)

32
64
128
512
1028

Image Size	Optimizers	Batch_Size	Fully-connected layer	Dropout
150	Adam	32		0.2

(d)

0.1
0.2
0.3
0.4
0.5

Image Size	Optimizers	Batch_Size	Fully-connected layer	Dropout
150	Adam	32	512	

(e)

APPENDIX I. (a) To find the best Image Size, the other parameters were initialized randomly except for the Batch_Size, the default value was used. (b) For the model to determine the optimal optimizer value, the best image size from a was applied, additionally taking the other hyperparameters' initial values from step a. (c) To find the best value for the Batch_Size hyperparameter by adapting the best values for each Image Size from a and the best optimizer from step b. (d) To obtain the appropriate Fully connected layer value, the best hyperparameter values from a, b, and c were used. (e) The best hyperparameters values from a, b, c, and d were used to find the best dropout value for the model.