




Article

Hybrid Facial Emotion Recognition Using CNN-Based Features

H. M. Shahzad ^{1,2} , Sohail Masood Bhatti ^{1,2}, Arfan Jaffar ^{1,2}, Sheeraz Akram ^{1,2,3,*} , Mousa Alhajlah ⁴
and Awais Mahmood ⁴ 

¹ Faculty of Computer Science and Information Technology, The Superior University, Lahore 54000, Pakistan; shahzad.dar@superior.edu.pk (H.M.S.)

² Intelligent Data Visual Computing Research (IDVCR), Lahore 54000, Pakistan

³ Information Systems Department, College of Computer and Information Sciences, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh 12571, Saudi Arabia

⁴ Computer Science and Information Systems Department, Applied Computer Science College, King Saud University, Riyadh 12571, Saudi Arabia

* Correspondence: sAkram@imamu.edu.sa

Abstract: In computer vision, the convolutional neural network (CNN) is a very popular model used for emotion recognition. It has been successfully applied to detect various objects in digital images with remarkable accuracy. In this paper, we extracted learned features from a pre-trained CNN and evaluated different machine learning (ML) algorithms to perform classification. Our research looks at the impact of replacing the standard SoftMax classifier with other ML algorithms by applying them to the FC6, FC7, and FC8 layers of Deep Convolutional Neural Networks (DCNNs). Experiments were conducted on two well-known CNN architectures, AlexNet and VGG-16, using a dataset of masked facial expressions (MLF-W-FER dataset). The results of our experiments demonstrate that Support Vector Machine (SVM) and Ensemble classifiers outperform the SoftMax classifier on both AlexNet and VGG-16 architectures. These algorithms were able to achieve improved accuracy of between 7% and 9% on each layer, suggesting that replacing the classifier in each layer of a DCNN with SVM or ensemble classifiers can be an efficient method for enhancing image classification performance. Overall, our research demonstrates the potential for combining the strengths of CNNs and other machine learning (ML) algorithms to achieve better results in emotion recognition tasks. By extracting learned features from pre-trained CNNs and applying a variety of classifiers, we provide a framework for investigating alternative methods to improve the accuracy of image classification.

Keywords: deep learning; facial emotions recognition; facial expression; convolution neural network; machine learning



Citation: Shahzad, H.M.; Bhatti, S.M.; Jaffar, A.; Akram, S.; Alhajlah, M.; Mahmood, A. Hybrid Facial Emotion Recognition Using CNN-Based Features. *Appl. Sci.* **2023**, *13*, 5572. <https://doi.org/10.3390/app13095572>

Academic Editor: Yu-Dong Zhang

Received: 8 March 2023

Revised: 21 April 2023

Accepted: 21 April 2023

Published: 30 April 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Emotion recognition has been studied by researchers in various fields, such as psychology, sociology, health care, entertainment, advertisement, education, robotics, and computer science [1]. Emotion recognition systems can be used in many applications, such as human-computer interaction, intelligent tutoring systems, surveillance, the psychological state of patients, and lie detection [2]. The ability to recognize facial expressions under masks has become increasingly important for creating realistic and engaging augmented reality (AR) and virtual reality (VR) experiences during the COVID-19 pandemic [3,4]. It is a challenging problem due to the many variables involved, such as facial features, lighting, and head pose. There are many approaches to solving this problem, but most of them require a substantial amount of training data [5].

Wearing face masks that cover the mouth and nose during a worldwide pandemic is strongly advised to reduce the risk of infection. This precautionary approach, though, might significantly impact how people interact with one another: only the highest parts

of the face, the eyes, and the forehead are visible while one is wearing a face mask, which raises the question of whether doing so would make it harder to identify emotions [6,7].

A study has shown that people wearing masks are less likely to be accurately identified because, with a hidden face, certain emotions like happiness, sadness, anger, etc. Due to this, participants are less accurate at identifying emotions when the person is wearing a mask [8]. In addition, the use of a face mask has also impaired people's ability to understand the gestures of others, which can be important for understanding the speaker's intentions [9]. According to the author, this could be because masks conceal the lower half of the face, which is where most of the crucial facial cues are found. Comparing masked targets with unmasked faces revealed a decline in emotion recognition accuracy. Target accuracy was significantly lower for masked faces (48.9 percent) than for unmasked faces (69.9 percent) [10].

The paper is structured into eight sections. Section 1 provides an introduction to the research problem and the paper's contributions. Section 2 presents a comprehensive literature review of the related work. In Section 3, the proposed solution is explained in detail. Section 4 discusses the implications of the pre-trained model. Section 5 provides a detailed explanation of the feature extraction process from the hidden layers in the neural network. Section 6 examines the various classifiers used in machine learning. Section 7 provides details of the dataset used in the experiments, MLF-W-FER, and explains the experiments conducted on the FC6, FC7, and FC8 layers of the VGG-16 and AlexNet architectures. Finally, the conclusions drawn from the experiments are presented in Section 8.

2. Related Work

In the past, this question has been addressed by several studies showing that the recognition of emotions expressed through the face is more difficult when the mouth is covered. A recent study suggested that this result was decreased because the mouth is responsible for most of the facial expressions of emotions and, therefore, when it is covered, there is less information to process. When most of the face is covered by a mask, the computer has a harder time understanding the facial expressions of the person [11,12]. Machine learning (ML) algorithms can be used to automatically detect and classify facial expressions. These algorithms typically use a combination of features extracted from the face, such as the shape of the eyebrow, the size of the eye, and the position of the mouth. The accuracy of these algorithms varies depending on the data set used for training, but generally, they can achieve good accuracy [13–15].

Some common ML methods used for emotion recognition include support vector machines (SVM), Linear Discriminant Analysis (LDA), k-nearest neighbors (k-NN), and decision trees (DT). These methods can be used to learn models from data that can be used to recognize facial emotions. Neural Network also have an important impact on this area of research [16]. Deep learning architectures are like special tools that can be used to achieve specific goals. Each architecture has its own advantages and disadvantages, so it is important to select the right one for the task at hand [17]. CNN is very effective for emotion recognition tasks [18]. They can extract features from input images, and then use these features to train a classifier. Once the classifier is trained, it can be used to recognize images. The advantage of using a CNN is that it is able to learn complex patterns in input images. This is important because it enables CNN to recognize images [19–24].

Recently, transfer learning methods have been widely used for prediction and classification. Transfer learning is a type of ML where knowledge gained from solving one problem is applied to a different but related problem. It is also an optimization technique that reduces the amount of training data required to solve a new ML problem [25]. The basic goal of transfer learning is to use the skills learned from one problem's resolution to aid in the solution of a related problem. This is especially helpful if you only have a small amount of data at your disposal. Numerous applications of transfer learning, including image classification, natural language processing, and even reinforcement learning, have been demonstrated to be successful [26].

Emotion recognition under a mask is difficult for ML algorithms [27] because there is often a lot of noise in the data that can make it difficult for computers to accurately recognize emotions. Additionally, it is difficult to accurately predict facial emotions when most of the face is covered by a mask [28]. As a result, the accuracy of emotion recognition decreases when compared with datasets that do not cover faces with masks.

There are various reasons for this decreased accuracy, including the loss of important facial features under the mask, such as the mouth and eyes. In addition, the mask can distort the shape of the face, making it difficult for the algorithm to correctly identify facial features. This can impact the accuracy of the algorithms, as they are not able to see the full range of emotions that people are expressing. This can be a problem in applications such as security, healthcare, advertisement, education, and robotics where facial recognition is used to identify people's expressions. It is possible to improve the accuracy of emotion recognition under a mask by using a deep learning approach. One approach is to use a deep learning algorithm to learn to extract features from images and apply a different regression technique of ML to predict face-masked expression.

This paper presents a contribution to the field of computer vision by exploring the impact of replacing the SoftMax classifier with alternative machine learning algorithms on the learned features extracted from DCNNs. Specifically, the study examines the efficacy of SVM and Ensemble classifiers on the FC6, FC7, and FC8 layers of AlexNet and VGG-16 architectures using a masked facial expression dataset. The results demonstrate that these alternative classifiers outperform the SoftMax classifier by achieving an accuracy improvement of 7% to 9% on each layer, highlighting the potential for combining the strengths of CNNs and other ML algorithms to enhance image recognition performance. Overall, this paper provides a framework for future investigations into alternative methods of improving image classification accuracy by leveraging learned features from pre-trained CNNs and applying a variety of classifiers.

3. Proposed Method

The initial stage of our system involves inputting images and performing feature extraction. This process involves utilizing pre-trained neural networks, specifically the AlexNet and VGG-16 models. The extracted features are obtained from the last three layers of this model. Activation Rectified Linear Unit (ReLU) is used in FC6 and FC7 layers, while SoftMax activation is used in FC8 as the classifier. The image representation that results from the concatenation of three fully connected layer features has 9192 dimensions features. This paper proposes a feature-based image classification technique to improve the accuracy of deep CNN models on masked facial expression data. The features from all images of face-masked facial expressions have been stored on each layer in a CSV file. Different classifiers have been used for image classification based on these features.

In the classifier phase, different classifiers have been applied, e.g., discriminant analysis, decision tree, support vector machine (SVM), k-nearest neighbor (KNN), and ensemble classifiers. In this paper, it is suggested to find which classifier has the highest accuracy on the masked facial expression dataset on each fully connected layer by using the two original pre-trained CNN models.

The proposed methodology for feature extraction and categorization is illustrated in Figure 1, which provides a high-level overview of the entire process. Furthermore, the detailed feature extraction and classification process is presented below.

1. Load the dataset (MLF-W-FER dataset).
2. Load pre-trained CNN (on AlexNet and VGG-16 model).
3. Divide the image sets into training and testing data (70% for training and 30% for testing).
4. Extract FC6, FC7, and FC8 features from the activation functions.
5. Train all classifiers (which are available in ML) on the MLF-FER-M dataset in FC6, FC7, and FC8.
6. Predict the class of the masked dataset for the test set using the trained classifier. e.g., positive, negative, and neutral.

7. Show the average accuracy of each classifier.
8. Compile the findings using a confusion matrix of the highest accuracy.

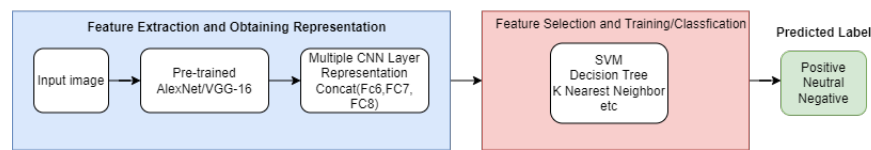


Figure 1. Alexnet and VGG16 feature extraction and classifier application for masked datasets.

3.1. Alexnet Architecture

The ImageNet database is used to train CNN, known as AlexNet, on more than a million photos. The 48-layer network can categorize photos into 1000 different items. A 224×224 pixel image in the ImageNet style serves as the network's input. A 1000-dimensional vector of class probabilities is the result. AlexNet was designed by Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton. It was first used to win the ImageNet Large Scale Visual Recognition Challenge in 2012. AlexNet is similar to the VGG network but is much lighter and faster. It is also the first network to use the ReLU activation function [29].

3.2. VGG-16 Architecture

The CNN VGG-16 has 16 layers in total. The Visual Geometry Group (VGG) at Oxford was responsible for its initial development. In ImageNet, a dataset of more than 14 million images was categorized into 1000 categories, and the model obtained a top-5 test accuracy of 92.7 percent. 13 convolutional layers make up VGG-16, along with 5 pooling layers and 3 fully connected layers. The succession of convolutional and pooling layers that make up the convolutional base is used to extract features from the input image. Following the transmission of these features, the fully linked layers use them to assign the image to one of the 1000 ImageNet classes [30,31]. In the next subsection, the pre-trained model is briefly described.

4. Pre-Trained Model

A pre-trained model of AlexNet or VGG-16 is useful for many computer vision applications as it has been trained on a large dataset and thus contains a lot of information about the images. Some potential applications for a pre-trained AlexNet or VGG-16 model include object detection, image classification, pattern recognition, image segmentation, texture recognition, object tracking, action recognition, and face recognition [32].

Through the process of transfer learning, previously learned information can be used to solve one problem and then applied to another. This method results in a considerable amount of information being contained in the model's weights, allowing for faster learning due to the efficient reuse of information [33].

The next subsection explains the feature extraction from AlexNet and VGG-16 architectures.

5. Feature Extraction

CNN, such as the AlexNet and VGG-16 designs, are both commonly employed for image classification applications. VGG-16 is made up of 16 convolutional layers as compared with AlexNet's 5 convolutional layers and 3 fully connected layers. VGG-16 uses the highest filter size of 4×4 , while AlexNet utilizes the smallest filter size of 3×3 [34].

AlexNet has a smaller capacity than VGG-16 but is faster to train. Both architectures extract features from images by convolving over the image with a set of learned filters. These features are then passed through a series of fully connected layers, which perform classification [35]. The facial image processing pipeline consists of two steps: facial feature extraction and detection. The pre-trained AlexNet model was used to extract features from facial image data at the output layer of the chain. We selected AlexNet as the CNN model for feature extraction based on a recent study, which compares the CNN model to five

well-known pre-trained models and shows that it appears to be good for age-invariant facial recognition systems [36].

The first layer extracts the input layer's average activation. The average activation of a layer determines how active the neurons in the layer are and is often used to determine the overall importance of a neuron in a layer. By using the average activation of a layer, we can determine which neurons are the most important for classification.

The second layer extracts the input layer's non-linearity. Non-linearity is determined by how much the input layer's activation changes when the input layer's weight is changed. This feature is important because it determines how well the neurons in the input layer map to the correct output category.

The third layer extracts the output layer's correlation with the input layer. The correlation between two layers is determined by how well the output of one layer predicts the input of another layer. This feature is important because it determines how well the neurons in the output layer map to the correct input category [37,38].

These features were extracted from VGG-16 and AlexNet, and different classifiers were applied to the masked face dataset of facial expressions. In the next subsection, we discuss the different classifiers that are used in our experiments.

6. Machine Learning Classifiers

ML classifiers are models that are used to predict the class of a data point. In ML, a classifier is an algorithm that maps an input data vector to a specific category or predicts the class of a data point. The input data can be continuous, such as an image, or discrete, such as an audio signal or text. Common classifiers include support vector machines, decision trees, and logistic regression, which is mentioned in Figure 2.

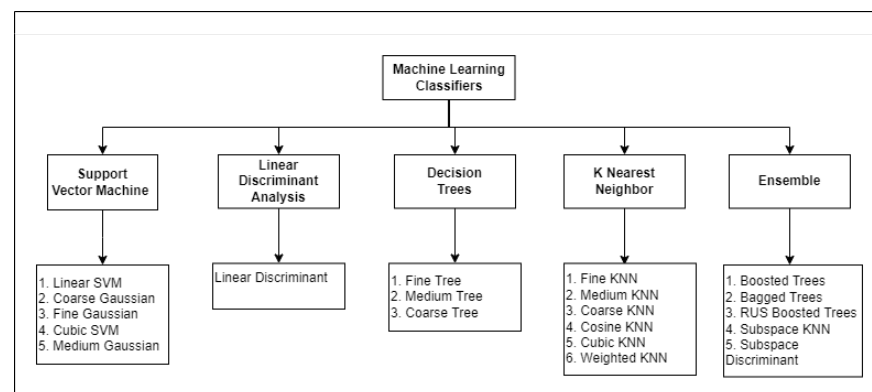


Figure 2. Alexnet and VGG16 feature extraction and classifier application for masked datasets.

6.1. Decision Tree

A decision tree classifier is an ML algorithm that can be used to predict the class of an instance based on the values of its features. A decision tree classifier works by repeatedly splitting the data into smaller groups based on a certain criterion until each group contains only instances with the same class label [39].

6.2. Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) is another supervised learning algorithm that can be used for both regression and classification tasks. It is a supervised learning algorithm, which means that it requires a training set of data in order to learn how to classify new items. The LDA classifier is a linear classifier, which means it divides data into two categories using a straight-line graph. This can be done by placing points along the x-axis of a coordinate grid and observing which diagonal lines cut through the points. From this information, LDA can calculate the distance or dissimilarity between two data samples [40].

6.3. Support Vector Machine

A support vector classifier works by mapping data to a high-dimensional space and then finding a hyperplane that separates the data. This hyperplane is defined by the support vectors, which are the points closest to it. The distance from the hyperplane is called the margin, and the aim is to maximize this margin [41].

The advantage of using a support vector classifier is that it can be more accurate than other methods, as it takes into account all of the data points when creating the hyperplane. It can also be used for non-linear classification tasks by using a kernel function.

6.4. K-Nearest Neighbor

A k-Nearest Neighbors (k-NN) algorithm is an unsupervised learning technique that can be applied to both classification and regression problems. It operates by locating the k nearest neighbors of a given data point and then predicting the label of the data point based on the labels of those neighbors. The algorithm operates by first figuring out how far apart each data point is from its closest neighbors. Then, it assigns a class label to each data point based on the class label that is most prevalent among its closest neighbors [42].

6.5. Ensemble

An ensemble classifier is a meta-classifier that combines the predictions of multiple base classifiers to form a more accurate final prediction. Ensemble classifiers are used in a variety of areas, including computer vision, speech recognition, and machine learning. Ensemble classifiers usually outperform single-model classifiers [43].

7. Experimental Setup and Results

7.1. Dataset

The experiment for this research is based on the M-MLFW-FER dataset. The M-LFW-FER [44] dataset is a collection of 11,038 images of faces wearing masks (6155 positive, 3988 neutral, and 895 negative). The statistics of the MLFW database are in Table 1.

Table 1. Statistic of mask count of MLFW database.

Positive	Neutral	Negative	Total
6155	3988	895	11,038

Each image is labeled with the identity of the person wearing a mask and the expression of the face. The M-MLFW-FER dataset is automatically created by detecting the location of the face and the mouth in each image, thereby cropping the image, and it includes only the face and the mask. The dataset is divided into three parts, which include neutral, positive, and negative expressions, as shown in Figure 3. In this research, the M-MLFW-FER dataset is used, which is available publicly.



Figure 3. Images from each emotion class in the M-LFW-FER dataset.

7.2. First Experiment: Transfer Learning from AlexNet for Extracting Features Using Different Classification

Instead of creating a CNN from scratch for this study, a pre-trained CNN was employed. Pre-trained networks can be used to extract features from a variety of image types. In this study, AlexNet and VGG-16, a network that was previously trained on ImageNet, are tested and applied to different classifiers to see which classifier is the best for the MLF-W-FER dataset of classification.

In this experiment, the performance is evaluated when transferring learning from the AlexNet network is used for feature selection and is applied to ML classifications, as shown in Figure 4.

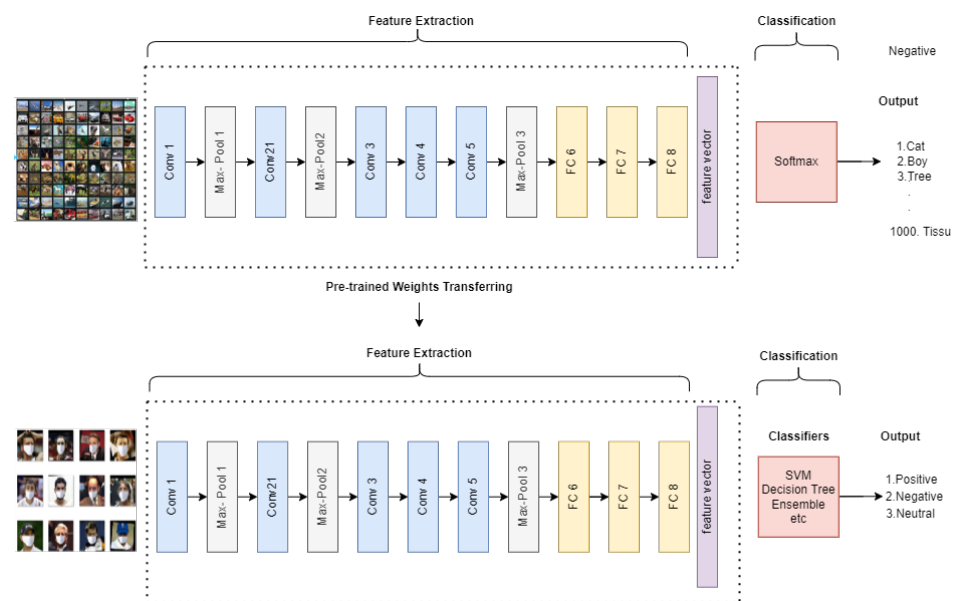


Figure 4. Transfer learning from AlexNet for extracting features and different classifiers techniques.

Table 2 depicts the results of an experiment wherein FC6 features were extracted from the AlexNet 240 architecture and subsequently subjected to various machine learning classifiers. Upon analyzing the outcomes, it was observed that the Subspace Discriminant classifier outperformed the others with a classification accuracy of 62.7%. The Quadratic SVM classifier stood second with a close accuracy score of 62.5%, while the Medium Gaussian SVM classifier secured the third position with a classification accuracy of 62.0%. These results suggest that the Subspace Discriminant classifier may be a more effective choice for the classification task at hand.

On the other hand, FC7 features were applied with different classifiers of ML. The results of the ML classifiers in the FC7 results are shown in Table 2. The result shows that Linear SVM achieved the highest accuracy of 62.0%, Quadratic SVM has the second best accuracy with 61.7%, followed by Medium Gaussian SVM with 61.6%. According to the results of FC8 features and performance comparison of ML classifiers, the best accuracy score was achieved by Subspace Discriminant with 62.6%, followed by Medium Gaussian SVM at 62.5%, and Quadratic SVM with 63.4%.

These tests show that Ensemble Classifiers and SVM classifier subclasses perform well on the pre-trained model of AlexNet features and that using different classifiers increased the accuracy of the masked dataset on each layer of FC6, FC7, and FC8. The classification results for the FC6, FC7, and FC8 layers of the AlexNet are mentioned in Table 2.

We obtained features from AlexNet and applied the Fitcecoc function to get the accuracies on each layer, FC6, FC7, and FC8. The Fitcecoc function returns full, trained, and error-correcting output codes (ECOC). ECOC [45] is a multiclass classifier designed for large training data sets. By using the Fitcecoc function in AlexNet, the accuracy of the

Fc6, Fc7, and Fc8 layers is 55% percent, 55.35% percent, and 55.64% percent, respectively, as shown in Table 3.

Table 2. Performance comparison of ML Algorithm Classification of MLF-W-FER Dataset on AlexNet.

Classifier	Classifier Type	FC6 Accuracy	FC7 Accuracy	FC8 Accuracy
Decision Tree	Fine Tree	55.5%	52.0%	53.8%
	Medium Tree	55.8%	54.4%	56.4%
	Coarse Tree	56.2%	53.8%	56.8%
Discriminant Analysis	Linear Discriminant	59.5%	48.4%	59.5%
Support Vector Machines (SVM)	Linear SVM	61.9%	62.0%	62.2%
	Quadratic SVM	62.5%	61.7%	63.4%
	Cubic SVM	59.6%	59.8%	60.8%
	Fine Gaussian SVM	55.8%	49.8%	55.9%
	Medium Gaussian SVM	62.0%	61.6%	62.5%
	Coarse Gaussian SVM	59.5%	59.9%	59.85
K Nearest Neighbor (KNN)	Fine KNN	50.3%	50.5%	50.9%
	Medium KNN	56.0%	54.5%	56.2%
	Coarse KNN	58.45%	57.0%	57.7%
	Cosine KNN	57.0%	55.5%	56.7%
	Cubic KNN	53.2%	54.4%	54.4%
	Weighted KNN	55.7%	55.4%	56.1%
Ensemble Classifiers	Boosted Trees	59.0%	57.2%	58.5%
	Bagged Trees	56.8%	56.2%	57.0%
	Subspace Discriminant	62.7%	59.7%	62.6%
	Subspace KNN	50.6%	50.8%	50.7%
	RUS Boosted Trees	44.4%	46.7%	44.6%

Table 3. Accuracy AlexNet pre-trained model on FC6, FC7, and FC8 on the masked dataset.

Alexnet	Accuracy
FC6	55.00%
FC6	55.35%
FC8	55.64%

We applied different classifiers to FC6, FC7, and FC8 individually. The results show that Subspace Discriminant, Linear SVM, and Quadratic SVM are the best methods for classification in FC6, FC7, and FC8, respectively, as shown in Table 4.

Table 4. Improved Accuracy of AlexNet pre-trained model on FC6, FC7, and FC8 with ML classifiers on the MLF-W-FER dataset.

AlexNet	Classifier	Accuracy
FC6	Subspace Discriminant	62.7%
FC7	Linear SVM	62.0%
FC8	Quadratic SVM	63.4%

Furthermore, the best accuracies of classifiers on each layer of the AlexNet model are mentioned in the confusion matrix (see Figure 5).

FC6				FC7				FC8			
Quadratic SVM (65.5%)				Linear SVM (62.5%)				Linear SVM (61.2%)			
Negative	87	238	278	Negative	0	345	128	Negative	28	256	319
Neutral	94	1308	1232	Neutral	0	1912	1161	Neutral	18	1345	1271
Positive	114	779	3194	Positive	0	1020	2628	Positive	19	801	3267
	Negative	Neutral	Positive		Negative	Neutral	Positive		Negative	Neutral	Positive

Figure 5. Best classifiers for AlexNet on FC6, FC7, and FC8 in the Confusion Matrix.

In Figure 5, the confusion matrix of FC7 Linear SVM was not able to predict the negative class of the MLF-W-FER dataset. That is why we selected the Quadratic SVM with the second highest accuracy (61.7%) because it is predicting all three classes as positive, negative, and neutral. The Linear SVM has been replaced with Quadratic SVM in Figure 6.

FC6				FC7				FC8			
Subspace Discriminant (62.7%)				Quadratic SVM (61.7%)				Quadratic SVM (63.4%)			
Negative	87	238	278	Negative	34	359	210	Negative	28	256	319
Neutral	94	1308	1232	Neutral	25	1918	1130	Neutral	18	1345	1271
Positive	114	779	3194	Positive	27	1054	2567	Positive	19	801	3267
	Negative	Neutral	Positive		Negative	Neutral	Positive		Negative	Neutral	Positive

Figure 6. Best classifiers for AlexNet on FC6, FC7, and FC8 in the Confusion Matrix.

7.3. Second Experiment: Transfer Learning from VGG-16 for Extracting Features Using Different Classification

In the second experiment, as depicted in Figure 7, the effectiveness of feature selection has been assessed using transfer learning from the VGG-16 network in conjunction with several classifications. According to the results in Table 4 for the FC6 features applied to the various classifiers, Quadratic SVM has a better classification accuracy of 65.5%, followed by Linear SVM with 65.3% and Cubic SVM with 64.4%.

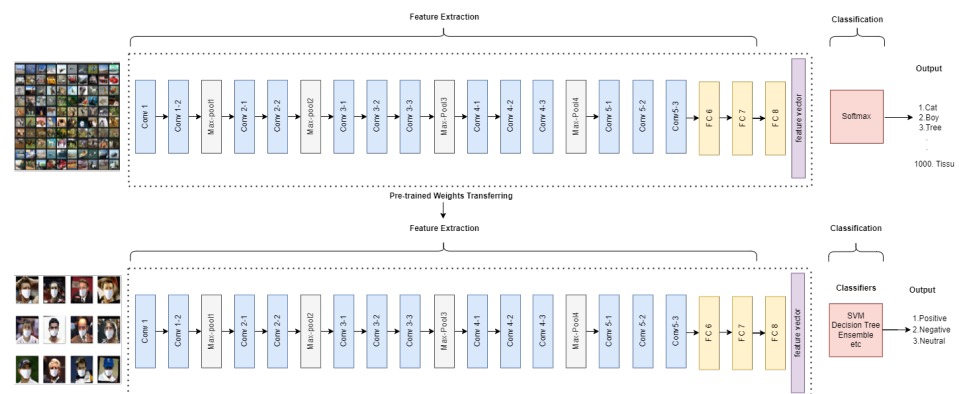


Figure 7. Transfer learning from VGG-16 for extracting features and different classifiers techniques.

The best accuracy score has been achieved by Linear SVM with a score of 62.0%, followed by Quadratic SVM with a score of 61.7%, and Medium Gaussian SVM with a score of 61.6%, according to the results of FC7 features and performance comparison of classifiers. According to the results of the FC8 features and performance comparison of classifiers,

Linear SVM has the highest accuracy score (62.0%), followed by Medium Gaussian SVM at 62.5 percent and Quadratic SVM at 63.4 percent, as shown in Table 5.

Table 5. Performance comparison of ML Algorithm Classification of MLF-W-FER Dataset on VGG-16.

Classifier	Classifier Type	FC6 Accuracy	FC7 Accuracy	FC8 Accuracy
Decision Tree	Fine Tree	53.6%	54.6%	54.2%
	Medium Tree	56.7%	55.7%	56.2%
	Coarse Tree	56.4%	56.2%	56.0%
Discriminant Analysis	Linear Discriminant	49.6%	51.4%	57.8%
Support Vector Machines (SVM)	Linear SVM	65.3%	62.5%	61.2%
	Quadratic SVM	65.5%	64.0%	59.8%
	Cubic SVM	64.4%	59.8%	57.4%
	Fine Gaussian SVM	55.8%	55.8%	55.8%
	Medium Gaussian SVM	64.2%	61.5%	60.3%
	Coarse Gaussian SVM	62.0%	59.8%	58.4
K Nearest Neighbor (KNN)	Fine KNN	52.8%	51.5%	50.9%
	Medium KNN	58.4%	57.2%	56.4%
	Coarse KNN	59.0%	58.3%	57.8%
	Cosine KNN	59.3%	56.3%	55.7%
	Cubic KNN	58.6%	57.2%	56.4%
	Weighted KNN	58.5%	57.0%	56.2%
Ensemble Classifiers	Boosted Trees	59.0%	58.4%	58.6%
	Bagged Trees	58.7%	57.8%	57.7%
	Subspace Discriminant	62.1%	58.0%	61.0%
	Subspace KNN	50.4%	50.8%	51.2%
	RUS Boosted Trees	48.3%	47.6%	47.0%

In this experiment, the features obtained from VGG-16 were fed into the Fitcecof function. As a result, the accuracy of each layer (FC6, FC7, and FC8) was 56.85%, 55.19%, and 56.73%, respectively, as shown in Table 6.

Table 6. Accuracy of VGG-16 pre-trained model on FC6, FC7, and FC8 on the masked dataset.

VGG-16	Accuracy
FC6	56.85%
FC6	55.19%
FC8	56.73%

Further, we applied ML classifiers on FC6, FC7, and FC8 individually to classify the masked dataset. The results show that Subspace Discriminant, Linear SVM, and Quadratic SVM are the best methods for classification in FC6, FC7, and FC8, respectively, in the masked dataset, as shown in Table 7. The accuracy of FC6, FC7, and FC8 was 65.5%, 62.5%, and 61.2%, respectively.

Table 7. Improved accuracy of VGG-16 pre-trained model on FC6, FC7, and FC8 with ML classifiers on the MLF-W-FER dataset.

VGG-16	Classifier	Accuracy
FC6	Subspace Discriminant	65.5%
FC7	Linear SVM	62.5%
FC8	Quadratic SVM	61.2%

Furthermore, the best accuracies of classifiers on each layer of the VGG-16 model are mentioned in the confusion matrix (see Figure 8).

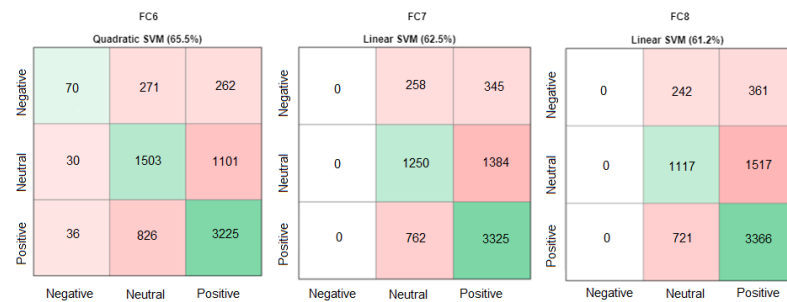


Figure 8. Best classifiers for VGG-16 on FC6, FC7, and FC8 in the Confusion Matrix.

Figure 8 shows that the Linear SVM of FC7 and FC8 was not the optimal classifier to use because it was unable to predict the negative classes. The second-highest accuracies in FC6 and FC7 were selected because they predicted all three classes of the MLF-W dataset.

We replaced Linear SVM (62.5%) with Quadratic SVM (62.4%) in FC7 and Linear SVM (61.2%) with Median Gaussian SVM (60.3%) in FC8, respectively, which is shown in Figure 9.

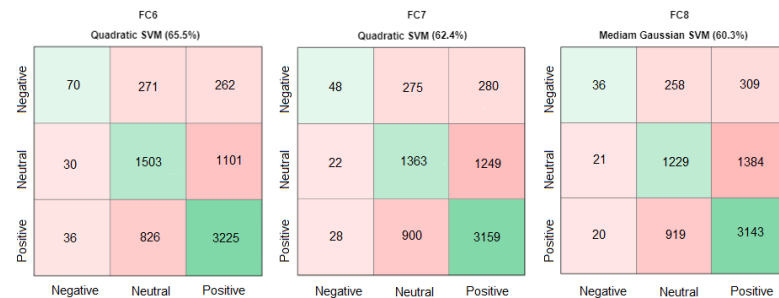


Figure 9. Best classifiers for VGG-16 on FC6, FC7, and FC8 in the Confusion Matrix.

In the confusion matrix, results showed that negative expressions are detected with low accuracy as compared with positive and neutral expressions on the M-MLFW-FER dataset. It is just because infrequent classes are predicted with low accuracy most of the time. The negative expression in the dataset is the lowest among other datasets, e.g., positive and negative expressions. In this paper, it is assessed how well the SVM classifier and its sub-classes perform on the pre-trained model of VGG-16 features and how applying different classifiers improves the accuracy of the masked dataset on each layer of FC6, FC7, and FC8.

Additionally, we have conducted a comparative analysis of our method's accuracy with the results obtained by different authors who employed the VGG-16 and VGG-19 architectures on the MLF-W-FER dataset. Their reported accuracies were 55.6% [46] and 53.54% [47], respectively. By comparing these results, our method has exhibited superior accuracy.

8. Conclusions

CNN AlexNet and VGG-16 are the two most popular deep learning models used for extracting features. After feature extraction, different classifiers were applied to FC6, FC7, and FC8 to find better prediction performance on masked facial expressions. We evaluated MLF-W-FER overall accuracy between 55.0% and 56.73% on AlexNet and VGG-16 pre-trained models. After extracting FC6, FC7, and FC8, we applied ML algorithms as classifiers on each layer of FC6, FC7, and FC8. We found that SVM improves the accuracy of the MLF-W dataset and outclasses other classifiers on ML, especially SoftMax. The experimental results in Table 8 show that the proposed technique improves accuracy by 4% to 9% on the M-MLFW-FER dataset on each layer of FC6, FC7, and FC8.

Table 8. Overall performance on AlexNet and VGG-16 Feature Extraction and applied ML classifiers on Masked Dataset.

AlexNet Accuracy	Accuracy	With Other Classifiers	Accuracy
FC6	55%	Subspace Discriminant	62.7%
FC7	55.35%	Quadratic SVM	61.7%
FC8	55.64%	Quadratic SVM	63.4%
VGG-16 Accuracy	Accuracy	With Other Classifiers	Accuracy
FC6	56.85%	Quadratic SVM	65.5%
FC7	55.19%	Quadratic SVM	62.4%
FC8	56.73%	Median Gaussian SVM	60.3%

Author Contributions: Conceptualization, H.M.S., S.M.B. and A.J.; data curation, S.M.B.; formal analysis, H.M.S., M.A. and A.M.; funding acquisition, M.A. and A.M.; investigation, H.M.S., S.M.B., A.J., S.A., M.A. and A.M.; methodology, H.M.S., S.M.B., M.A. and A.M.; project administration, S.M.B.; resources, A.J. and S.A.; software, H.M.S.; supervision, S.M.B.; validation, A.J., S.A. and M.A.; visualization, H.M.S., S.M.B., A.J., S.A., M.A. and A.M.; writing—original draft, H.M.S.; writing—review and editing, H.M.S., S.M.B., A.J., S.A., M.A. and A.M. All authors have read and agreed to the published version of the manuscript.

Funding: Researchers Supporting Project number (RSP2022R458), King Saud University, Riyadh, Saudi Arabia.

Institutional Review Board Statement: Not Applicable.

Informed Consent Statement: Not Applicable.

Data Availability Statement: The data describe in this article are openly available in github at <https://github.com/KDDI-AI-Center/LFW-emotion-dataset/blob/main/README.md>, accessed on 20 April 2023.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Leo, M.; Carcagni, P.; Mazzeo, P.L.; Spagnolo, P.; Cazzato, D.; Distant, C. Analysis of facial information for healthcare applications: A survey on computer vision-based approaches. *Information* **2020**, *11*, 128. [\[CrossRef\]](#)
2. Pal, S.; Mukhopadhyay, S.; Suryadevara, N. Development and progress in sensors and technologies for human emotion recognition. *Sensors* **2021**, *21*, 5554. [\[CrossRef\]](#)
3. Fard, A.P.; Mahoor, M.H. Ad-corre: Adaptive correlation-based loss for facial expression recognition in the wild. *IEEE Access* **2022**, *10*, 26756–26768. [\[CrossRef\]](#)
4. Minaee, S.; Liang, X.; Yan, S. Modern Augmented Reality: Applications, Trends, and Future Directions. *arXiv* **2022**, arXiv:2202.09450.
5. Ko, B.C. A brief review of facial emotion recognition based on visual information. *Sensors* **2018**, *18*, 401. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Green, J.; Staff, L.; Bromley, P.; Jones, L.; Petty, J. The implications of face masks for babies and families during the COVID-19 pandemic: A discussion paper. *J. Neonatal Nurs.* **2021**, *27*, 21–25. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Giovanelli, E.; Valzoler, C.; Gessa, E.; Todeschini, M.; Pavani, F. Unmasking the difficulty of listening to talkers with masks: Lessons from the COVID-19 pandemic. *i-Percept.* **2021**, *12*, 2041669521998393. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Mheidly, N.; Fares, M.Y.; Zalzale, H.; Fares, J. Effect of face masks on interpersonal communication during the COVID-19 pandemic. *Front. Public Health* **2020**, *8*, 582191. [\[CrossRef\]](#)
9. Grahlow, M.; Rupp, C.I.; Derntl, B. The impact of face masks on emotion recognition performance and perception of threat. *PLoS ONE* **2020**, *17*, e0262840. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Grundmann, F.; Epstude, K.; Scheibe, S. Face masks reduce emotion-recognition accuracy and perceived closeness. *PLoS ONE* **2021**, *16*, e0249792. [\[CrossRef\]](#)
11. Sarker, I.H. Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions. *SN Comput. Sci.* **2021**, *2*, 420. [\[CrossRef\]](#)
12. Sultan Zia, M.; Hussain, M.; Arfan Jaffar, M. A novel spontaneous facial expression recognition using dynamically weighted majority voting based ensemble classifier. *Multimed. Tools Appl.* **2018**, *77*, 25537–25567. [\[CrossRef\]](#)
13. Jiménez, A.A.; Muñoz, C.Q.G.; Márquez, F.P.G. Dirt and mud detection and diagnosis on a wind turbine blade employing guided waves and supervised learning classifiers. *Reliab. Eng. Syst. Saf.* **2019**, *184*, 2–12. [\[CrossRef\]](#)

14. Kim, J.H.; Poullose, A.; Han, D.S. The extensive usage of the facial image threshing machine for facial emotion recognition performance. *Sensors* **2021**, *21*, 2026. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Canal, F.Z.; Müller, T.R.; Matias, J.C.; Scotton, G.G.; de Sa Junior, A.R.; Pozzebon, E.; Sobieranski, A.C. A survey on facial emotion recognition techniques: A state-of-the-art literature review. *Inf. Sci.* **2022**, *582*, 593–617. [\[CrossRef\]](#)
16. Tian, C.; Ma, J.; Zhang, C.; Zhan, P.A. deep neural network model for short-term load forecast based on long short-term memory network and convolutional neural network. *Energies* **2018**, *11*, 3493. [\[CrossRef\]](#)
17. Bouktif, S.; Fiaz, A.; Ouni, A.; Serhani, M.A. Multi-sequence LSTM-RNN deep learning and metaheuristics for electric load forecasting. *Energies* **2020**, *13*, 391. [\[CrossRef\]](#)
18. Karnati, M.; Seal, A.; Bhattacharjee, D.; Yazidi, A.; Krejcar, O. Understanding Deep Learning Techniques for Recognition of Human Emotions using Facial Expressions: A Comprehensive Survey. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 5006631. [\[CrossRef\]](#)
19. Quan, Y.; Chen, Y.; Shao, Y.; Teng, H.; Xu, Y.; Ji, H. Image denoising using complex-valued deep CNN. *Pattern Recognit.* **2021**, *111*, 107639. [\[CrossRef\]](#)
20. Elngar, A.A.; Arafa, M.; Fathy, A.; Moustafa, B.; Mahmoud, O.; Shaban, M.; Fawzy, N. Image classification based on CNN: A survey. *J. Cybersecur. Inf. Manag.* **2021**, *6*, 18–50. [\[CrossRef\]](#)
21. Pathak, Y.; Shukla, P.K.; Tiwari, A.; Stalin, S.; Singh, S. Deep transfer learning based classification model for COVID-19 disease. *Irbm* **2022**, *43*, 87–92. [\[CrossRef\]](#)
22. Liu, M.; Li, B.; Zhang, W. Research on Small Acceptance Domain Text Detection Algorithm Based on Attention Mechanism and Hybrid Feature Pyramid. *Electronics* **2022**, *11*, 3559. [\[CrossRef\]](#)
23. Zaman, K.; Sun, Z.; Shah, S.M.; Shoaib, M.; Pei, L.; Hussain, A. Driver Emotions Recognition Based on Improved Faster R-CNN and Neural Architectural Search Network. *Symmetry* **2022**, *14*, 687. [\[CrossRef\]](#)
24. Hussain, T.; Iqbal, A.; Yang, B.; Hussain, A. Real time violence detection in surveillance videos using Convolutional Neural Networks. *Multimed. Tools Appl.* **2022**, *81*, 38151–38173.
25. Tammina, S. Transfer learning using vgg-16 with deep convolutional neural network for classifying images. *IJSRP* **2019**, *9*, 143–150. [\[CrossRef\]](#)
26. Naseer, A.; Rani, M.; Naz, S.; Razzak, M.I.; Imran, M.; Xu, G. Refining Parkinson's neurological disorder identification through deep transfer learning. *Neural Comput. Appl.* **2020**, *32*, 839–854. [\[CrossRef\]](#)
27. Ejaz, M.S.; Islam, M.R.; Sifatullah, M.; Sarker, A. Implementation of principal component analysis on masked and non-masked face recognition. In Proceedings of the 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), Dhaka, Bangladesh, 3–5 May 2019; pp. 1–5.
28. Genç, Ç.; Colley, A.; Löchtefeld, M.; Häkkinä, J. Face mask design to mitigate facial expression occlusion. In Proceedings of the 2020 ACM International Symposium on Wearable Computers, Virtual Event, 4 September 2020; pp. 40–44.
29. Peng, P.; Zhao, X.; Pan, X.; Ye, W. Gas classification using deep convolutional neural networks. *Sensors* **2018**, *18*, 157. [\[CrossRef\]](#)
30. Gupta, R.K.; Sahu, Y.; Kunhare, N.; Gupta, A.; Prakash, D. Deep learning based mathematical model for feature extraction to detect corona virus disease using chest x-ray images. *Int. J. Uncertain. Fuzziness -Knowl.-Based Syst.* **2021**, *29*, 921–947. [\[CrossRef\]](#)
31. Jaworek-Korjakowska, J.; Kleczek, P.; Gorgon, M. Melanoma thickness prediction based on convolutional neural network with VGG-19 model transfer learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, CA, USA, 16–17 June 2019.
32. Caron, M.; Bojanowski, P.; Joulin, A.; Douze, M. Deep clustering for unsupervised learning of visual features. In Proceedings of the European conference on computer vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 132–149.
33. Lu, J.; Behbood, V.; Hao, P.; Zuo, H.; Xue, S.; Zhang, G. Transfer learning using computational intelligence: A survey. *Knowl.-Based Syst.* **2015**, *80*, 14–23. [\[CrossRef\]](#)
34. Mohanakurup, V.; Parambil Gangadharan, S.M.; Goel, P.; Verma, D.; Alshehri, S.; Kashyap, R.; Malakhil, B. Breast cancer detection on histopathological images using a composite dilated Backbone Network. *Comput. Intell. Neurosci.* **2022**, *2022*, 8517706. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Guan, Q.; Wang, Y.; Ping, B.; Li, D.; Du, J.; Qin, Y.; Xiang, J. Deep convolutional neural network VGG-16 model for differential diagnosing of papillary thyroid carcinomas in cytological images: A pilot study. *J. Cancer* **2019**, *10*, 4876. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Samir, S.; Emary, E.; El-Sayed, K.; Onsi, H. Optimization of a pre-trained AlexNet model for detecting and localizing image forgeries. *Information* **2020**, *11*, 275. [\[CrossRef\]](#)
37. Hegde, R.B.; Prasad, K.; Hebbar, H.; Singh, B.M.K. Feature extraction using traditional image processing and convolutional neural network methods to classify white blood cells: A study. *Australas. Phys. Eng. Sci. Med.* **2019**, *42*, 627–638. [\[CrossRef\]](#)
38. Uddin, M.A.; Pathan, R.K.; Hossain, M.S.; Biswas, M. Gender and region detection from human voice using the three-layer feature extraction method with 1D CNN. *J. Inf. Telecommun.* **2022**, *6*, 27–42. [\[CrossRef\]](#)
39. Bhavsar, H.; Ganatra, A. A comparative study of training algorithms for supervised machine learning. *IJSCE* **2012**, *2*, 2231–2307.
40. Saranya, T.; Sridevi, S.; Deisy, C.; Chung, T.D.; Khan, M.A. Performance analysis of machine learning algorithms in intrusion detection system: A review. *Procedia Comput. Sci.* **2020**, *171*, 1251–1260. [\[CrossRef\]](#)
41. Bi, Q.; Goodman, K.E.; Kaminsky, J.; Lessler, J. What is machine learning? A primer for the epidemiologist. *Am. J. Epidemiol.* **2019**, *188*, 2222–2239. [\[CrossRef\]](#)
42. Kang, S. K-nearest neighbor learning with graph neural networks. *Mathematics* **2021**, *9*, 830. [\[CrossRef\]](#)

43. Ashraf, M.; Zaman, M.; Ahmed, M. An intelligent prediction system for educational data mining based on ensemble and filtering approaches. *Procedia Comput. Sci.* **2020**, *167*, 1471–1483. [[CrossRef](#)]
44. Wang, C.; Fang, H.; Zhong, Y.; Deng, W. Mlfw: A database for face recognition on masked faces. In Proceedings of the Biometric Recognition: 16th Chinese Conference, CCBR 2022, Beijing, China, 11–13 November 2022.
45. Zou, J.Y.; Sun, M.X.; Liu, K.H.; Wu, Q.Q. The design of dynamic ensemble selection strategy for the error-correcting output codes family. *Inf. Sci.* **2021**, *571*, 1–23. [[CrossRef](#)]
46. Yang, B.; Wu, J.; Hattori, G. Facial expression recognition with the advent of face masks. In Proceedings of the 19th International Conference on Mobile and Ubiquitous Multimedia, Essen, Germany, 22 November 2020.
47. Shahzad, H.M.; Bhatti, S.M.; Jaffar, A.; Rashid, M. A Multi-Modal Deep Learning Approach for Emotion Recognition. *Intell. At. Soft Comput.* **2023**, *36*, 1561–1570. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.