Araştırma Makalesi https://doi.org/10.46810/tdfd.1022507 Research Article

BİNGÖL ÜNİVERSİTESİ

Türk Doğa ve Fen Dergisi
Turkish Journal of Nature and Science

www.dergipark.gov.tr/tdfd

Proposal of New Dataset for Child Face Expression Recognition and Comparison of Deep Learning Models on The Proposed Dataset

İrem SAYIN[1], Bekir AKSOY[2*]

[1] Yıldız Technical University, Machine Faculty, Mechatronics Engineering Department, İstanbul, Türkiye
[2] Isparta University of Applied Sciences, Technology Faculty, Mechatronics Engineering Department, Isparta, Türkiye
İrem SAYIN ORCID No: 0000-0002-0627-8308
Bekir AKSOY ORCID No: 0000-0001-8052-9411

**Corresponding author: bekiraksoy@isparta.edu.tr*

(Received: 12.11.2022, Accepted: 4.01.2023, Online Publication: 27.03.2023)

Keywords
Child Facial Expression Recognition, Facial Expression Recognition, Deep Learning, Transfer learning

Abstract: With the developing technology, smart systems have started to take place in our daily lives. Accordingly, it is very important for the systems that will actively participate in social life to adapt to social life properly. One of the most important steps of adapting to social life is communication. Facial expressions are one of the most important parts of communication that usually supports verbal communication. For this reason, many studies have been carried out on identifying facial expressions. The vast majority of these studies were carried out using datasets containing only adult faces. Conducting studies that do not involve the elderly and children may lead to the creation and development of highly biased smart systems. Therefore, this article focuses on detecting children's facial expressions. In order to detect facial expressions in children, a data set was prepared with images collected from search engines using keywords. In the study, VGG16, ResNet50, DenseNet121, InceptionV3, InceptionResNetV2 and Xception artificial intelligence models were used on the data set prepared using transfer learning methods, and the success of the models at the end of the training was evaluated and compared. Obtained results were evaluated according to performance evaluation metrics. According to the evaluation results, the best result was obtained with the InceptionV3 model with an accuracy rate of 76.3% and an F1 score of 0.76.

Çocuklarda Yüz İfadesi Tanımlama için Yeni Veri Seti Önerilmesi ve Veri Seti Üzerinde Derin Öğrenme Modellerinin Karşılaştırılması

Anahtar Kelimeler
Çocuklarda Yüz İfadesi Tespiti, Yüz İfadesi Tanıma, Derin Öğrenme, Transfer Öğrenme

Öz: Gelişen teknoloji ile akıllı sistemler günlük hayatımızda yer edinmeye başlamıştır. Sosyal hayatta aktif olarak katılacak sistem ve teknolojilerin sosyal hayata uyum sağlamaları oldukça önemlidir. Sosyal hayata uyum sağlamanın en önemli adımlarından birisi iletişimdir. Yüz ifadeleri genellikle sözlü olarak gerçekleştirilen iletişimi destekleyen iletişimin oldukça önemli parçalarından biridir. Bu nedenle son zamanlarda oldukça popüler bir alan olmuş olan yüz ifadelerini tanımlama üzerinde pek çok çalışma gerçekleştirilmiştir. Gerçekleştirilen bu çalışmaların büyük bir çoğunluğu yalnızca yetişkin yüzlerinin içeren veri setleri kullanılarak gerçekleştirilmiştir. Yaşlı ve çocukları içermeyen çalışmaların yapılması oldukça yanlı sistemlerin oluşturulması ve geliştirilmesine neden olabilir. Bu nedenle bu makalede ihmal edilen gruplardan bir tanesi olan çocuklar yüzleri üzerinde bir çalışma gerçekleştirilmiştir. Çalışmada arama motorlarında belirlenmiş olan anahtar kelimeler kullanılarak çocuk yüz ifadelerini içeren bir veri set hazırlanmıştır. Çalışmada transfer öğrenme yöntemleri kullanılarak hazırlanan veri seti üzerinde VGG16, ResNet50, DenseNet121, InceptionV3, InceptionResNetV2 ve Xception yapay zeka modelleri kullanılarak eğitim sonucunda modellerinin başarısı değerlendirilerek karşılaştırılmıştır. Elde edilen sonuçlar performans değerlendirme metriklerine göre değerlendirilmiştir. Değerlendirme sonuçlarına göre en iyi sonuç %76,3 doğruluk oranı ve 0,76 F1 puanı ile InceptionV3 modeli ile elde edilmiştir.

1. INTRODUCTION

Humans are social beings by nature. People communicate with each other to convey information, meet their social needs, and share their thoughts and feelings. This communication is usually carried out by verbal communication. Where verbal communication is insufficient, human expressions come into play. Facial Expressions have a very important place in human communication [1]. Understanding human expressions allows to determine the emotional state of the person at that moment. Deeper meanings can be extracted by combining the person's expressions and the emotions they reflect with their verbal expressions [2].

With the developing technology, simple verbal communications have begun to be used to assign simple tasks between humans and intelligent systems. The relationship between technology and people, which is expected to develop further in the future and take more place in daily life, needs to be developed. Intelligent technologies that will enter our lives will need to detect facial expressions of people, make mood analyses and communicate at a higher level. For this reason, facial expression recognition has become a very popular field and a lot of researches have been done on it and different datasets have been developed for this purpose

When the studies on expression recognition are examined, it is seen that the majority of the studies and the prepared data sets are mostly composed of adult faces. It is seen that the faces of children and the elderly are rarely included in the studies. In the study of Howard et al. on the bias of machine learning algorithms on facial expression detection, they stated that the old and young age groups and ethnic minorities were not adequately represented in the studies and data sets [3]. They stated that this causes machine learning algorithms to be biased. Guo et al.'s study on elderly faces found that elderly people's expressions were less exaggerated compared to adults. In addition, they stated that the loss of elasticity of the skin and wrinkles due to aging cause difficulty in detecting expression [4]. Similarly, Houstis and Kiliaridis examined the differences between children and adults in their study [5]. According to their study, they found differences in the way adult and child faces express emotions. They stated that vertical features in facial emotional expressions were not developed in children.

As a result of this unbalanced distribution in the field, it is likely that the systems to be developed in the future (smart home systems, assistive robots, etc.) will develop a bias towards a single age, gender and ethnic group. For example, in the study of Brandao and Martim on pedestrian identification algorithms, they found that the miss rate of female and child pedestrians was two times higher than that of male and adult pedestrians in the most successful algorithm [6]. In addition to the bias in the studies, the majority of the datasets containing facial expressions also consist of adults (Extended Cohn-Kanade Dataset/CK+, FER-2013, AffectNet, The Japanese Female Facial Expression Dataset/JAFFE, CMU MultiPIE). In addition to the small number of datasets created for children, these datasets generally consist of images collected in controlled environments. The majority of studies on the detection of child facial expressions use NIMH-ChEFS (The NIMH Child Emotional Faces Picture Set) [7] or CAFE (The Child Affective Facial Expression Set) [8] datasets.

In this study, a chil facial expression data set was created by collecting images from search engines for children between the ages of 2 and 10. While creating the data set, facial expressions belonging to 7 classes were collected, including angry, happy, disgusted, sad, scared, surprised and neutural, considering the 6 basic facial expressions presented by Ekman et al. [9]. Using this data set, different deep learning models were trained with the transfer learning method. The most successful model was determined by comparing the precision, recall, accuracy and F1 score values of the models.

In the second part, the main studies on child facial expression recognition are examined. In the third section, brief information is given about the models and data set. In the fourth section, comparisons of models are given according to precision, recall, F1 score and accuracy values. In the fifth and last chapter, the conclusion and plans for future studies are given.

2. RELATED WORKS

In the study carried out, academic studies on emotion analysis of children's facial expressions using artificial intelligence methods were examined in detail and discussed in detail below.

Rao et al., (2020) classified and evaluated the datasets of children and adults using the same model they developed in their study and investigated whether there are significant differences between adult and child faces for face detection [10]. They used the CK+ and CAFFE datasets to compare adult and child facial expressions. They identified sixty-eight key points using facial landmarks to be used in expression detection. They aimed to determine the minimum number of points required for adult and child faces by comparing the accuracy results they obtained with the DNN model they created. They tested the datasets with the DNN model using different facial landmark points. When all sixty-eight points were used with the CAFE dataset, they achieved an accuracy rate of 66%. in addition, the accuracy decreased to 49% when only twenty-three eye-brow points were used, and to 41% when only twenty-one lip points were used. In the CK+ dataset, they achieved an accuracy rate of 87% for all 68 points, 83% for eye-brow points, and 83% for lip points. They stated that the results they obtained support the hypothesis that there are differences in the ability of children and adults to express emotions. Finally, by analyzing the results of the CK+ dataset and maintaining the obtained accuracy, they tried to reduce the required number of facial points and reduced the eye-brow points to ten and the lip points to eight. For the CAFE dataset, they stated that the removal of facial points greatly affected the accuracy rate. They stated that children cannot display strong expressions as adults, and therefore

the same features may not be appropriate for children and adults.

Leo et al. developed an SVM-based expression recognition system to be used with the Robokind R25 robot to interact with children with Autism Spectrum Disorder (ASD) in their study [11]. With the proposed system, they aimed to automatically manage the medical protocol aimed at improving the capacity of children affected by ASD to associate emotions with facial expressions. The protocol performed is based on the imitation of the emotion shown by the robot by the child. The emotion imitated by the child was detected and analyzed by the robot with the model developed in the study. In the first step of the developed model, after the face identification process, the HOG vectors of these images were extracted with the Directed Gradient Histogram algorithm. These extracted HOG vectors were given as input to the SVM classifier and they performed the classification process. The training of this developed model was carried out using a subset of the CK+ dataset. As a result of the training, they obtained an accuracy rate of 97.5% with the model CK+ data set.

Nagpal et al., (2019) stated that children's faces are not adequately represented in expression identification studies [12]. In their study, they proposed a new learning model called Mean Supervised Deep Boltzmann Machine/ms-DBM in order to classify children's facial expressions. In the model they proposed, they enabled the model to learn the distinguishing features by minimizing the intra-class variations and maximizing the inter-class variations according to the average feature vectors. In their study, they trained and tested the model not only on children's faces but also on adult faces. For this reason, in addition to the Radbound Faces dataset, which includes both adult and child expressions, they used CAFE emotion datasets consisting only of child expressions. In order to perform expression recognition on children's faces, a pre-training was carried out with the images of 1197 adult facial expressions in the Radbound Faces dataset. As a result of their study, they obtained a better result by approximately 3.3% than the models they compared with the msDBM model on the Radboud dataset with an accuracy rate of 75%. On the CAFE dataset, they achieved an accuracy rate of 48%, which was approximately 14% better than the other models. They stated that the reason why this ratio is very different from the Radboud Faces dataset is that the images of children in the Radboud dataset consist of adolescent children, while the images in the CAFE dataset consist of images of younger children (2-8) years old. In the classification of adult expressions, they achieved an accuracy rate of 85.9% with the msDBM model.

Witherow et al., (2019) suggested using the transfer learning method for facial expression recognition in children [13]. With the method they proposed, they offered a solution to the model education problem that emerged due to the datasets containing the child expressions, which are very few in number. They used the CNN model they designed in the study. After training the CNN model they developed with the CK+ dataset consisting of adult face images, they fine-tuned it using the CAFFE dataset. Using the model they created, they conducted training in five different combinations. These combinations are;

1. Combination: Training and testing the model with the CK+ dataset.
2. Combination: Training and testing the model with the CAFE dataset.
3. Combination: Testing the model trained with CK+ with the CAFE dataset.
4. Combination: Creating a model using the weights of the model trained with the CK+ dataset and testing the model by training with the CAFE dataset.
5. Combination: Fine-tuning the model trained with the CK+ dataset using the CAFE dataset.

As a result of all these trainings, they obtained test accuracy rates of approximately 93% for the first case, approximately 63% for the second case, approximately 46% for the third case, approximately 62% for the fourth case, and approximately 76% for the fifth case. According to the results obtained, the best accuracy rate for the CAFE dataset was obtained with the model trained on the CK+ dataset, fine-tuned and tested using the CAFE dataset.

Lopez-Rincon (2019) performed facial expression recognition in children using the NAO robot in their study [14]. A four-layer CNN model was created to be used in the study and the created CNN model was compared with the Facial Action Coding System (FACS) based AFFDEX SDK classifier. The hyperparameters of the created model were determined and adjusted according to the results of bayesian optimization and trained using a subset of the AffectNet dataset. As a result of the training, they achieved an accuracy rate of 68.4% on the AffectNet validation dataset. In order to increase the accuracy of the same model on the CAFE dataset, it has been fine-tuned by training the model with the NIHM-ChEF dataset.

3. DATASET AND MODELS

In this section, information about the data set prepared in the study is given. In addition to that, the training parameters of the models used for comparison are also presented in this section.

3.1. Data Set

The data set prepared in this study was created by collecting the images from various search engines with bulk download tools using keywords. The data set used consists of child facial expressions and includes a total of seven expression classes. Images that are similar and do not contain human faces among the collected images were deleted and cleaned. All images were reviewed by three adult individuals and reclassified according to dominant expression to ensure that the images obtained were classified correctly. The faces in the classified images were detected using HaarCascade, and the cut images were resized to 128 x 128. Some sample images of the data set are given in Figure 1.

Figure 1. Examples from the data set

A total of 6364 images were obtained. However, in order to prevent the imbalance between classes in the data set, the amount of data in the classes was reduced to be close to the number of data contained in the 'disgust expression' class, where the least image was obtained. As a result of this process, all classes were arranged and the total number of data was reduced to 4002. The dataset was divided into three as 80% training, 10% validation and 10% test dataset. The training, validation and test data set distributions for all classes are given in Table 1.

Table 1. Distribution of the classes in the Data Set in the training, validation and test sets

Class Name	Training Dataset	Validation Dataset	Test Dataset	Total Data
Angry	467	54	54	575
Disgusted	416	48	48	512
Happy	500	58	58	616
Neutural	455	52	52	559
Sad	451	52	52	555
Scared	483	56	56	595
Surprised	478	56	56	590
Total	3 250	376	376	4 002

Evaluation of the models was carried out using the test data set. Methods were used to increase the amount of data used. For this purpose, random rotation, horizontal flip, horizontal and vertical scrolling operations were performed on the data set.

3.2. Models

In this study, 6 deep learning models, namely VGG16, ResNet50, DenseNet121, InceptionV3, InceptionResNetV2 and Xception, which were previously trained on ImageNet dataset, were used in order to train by transfer learning. The models were trained by fine tuning through freezing the weights in some layers. The most successful configuration obtained with the model as a result of many different trainings is presented in the study. All models were trained using early stopping according to the validation loss rate. Using Modelcheckpoint, validation accuracy and validation loss were tracked during the training of the models, and the models with the lowest validation loss and the highest validation accuracy were recorded.

The VGG16 model was first used to train the VGG16 in the study. The last layers of the model were removed and replaced with Global average pooling layer, dropout layer with 0.7 drop rate and 7 unit Softmax layers. Layers added to VGG16 are given in Figure 2. The weights of all layers except the last six layers were frozen. It was trained for 67 epochs using 128 batch size, 10^{-4} learning rate and Adam optimizer.

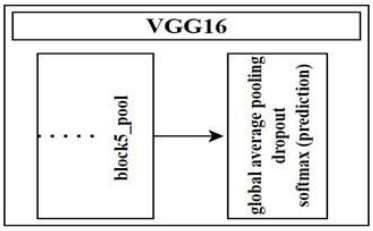

Figure 2. Layers added to the VGG16 model

Instead of the upper layers removed in the ResNet50 model, Global average pooling, dropout layer with 0.7 drop rate and 7 unit Softmax layer has been added as VGG16. Layers added to ResNet50 are given in Figure 3. Only the last 50 layers are trained using Batch size of 128, learning rate and Adam optimizer.

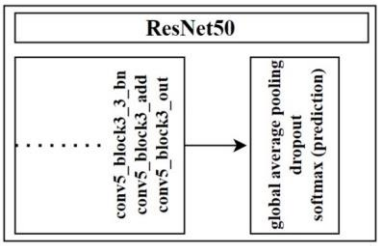

Figure 3. Layers added to the ResNet50 model

In the DenseNet121 model, the upper layers have been replaced with 0.7 dropout layer, global average pooling, 514 unit fully connected layer and 7 unit softmax layer. Layers added to DenseNet121 are displayed in Figure 4. 64 batch sizes were trained for 39 epochs using the Adam optimizer. When the validation loss plateus learning rate divided and decresed with factor 0.2 initial learning rate was determined as , and learning rate was reached at the end of the training.

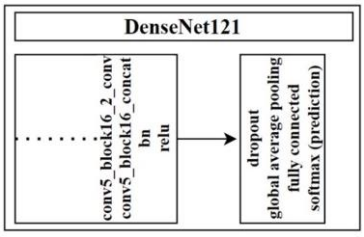

Figure 4. Layers added to the DenseNet121 model

In the InceptionV3 model, the upper layers have been replaced with global average pooling, 0.5 dropout layer, 1028 unit fully connected layer and 7 unit softmax layer. Layers added to InceptionV3 are given in Figure 5. Training was performed for 24 epochs using 64 batch size,

Adam optimizer. When the validation loss plateus learning rate divided and decresed with factor 0.2 initial learning rate was determined as , and learning rate was reached at the end of the training.

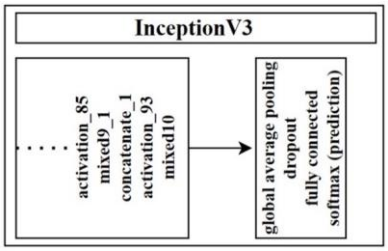

Figure 5. Layers added to the InceptionV3 model

In the InceptionResNetV2 model, the upper layers have been replaced with global average pooling and 7 softmax layers. Layers added to InceptionResNetV2 are presented in Figure 6. Training was carried out for 36 epochs using 64 batch size, Adam optimizer. When the validation loss plateus learning rate divided and decresed with factor 0.2 initial learning rate was determined as, a learning rate of was reached at the end of the training.

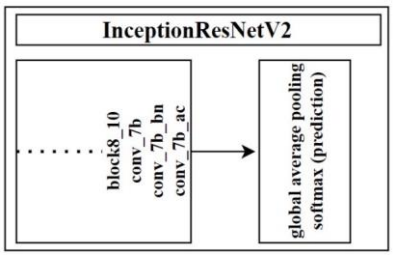

Figure 6. Layers added to the InceptionResNetV2 model

In the Xception model, the upper layers are replaced with global average pooling, 0.5 dropout layer, and 7 unit softmax layer. The layers added to Xception are given in Figure 7. Training was carried out for 36 epochs using 64 batch size, Adam optimizer. When the validation loss plateus learning rate divided and decresed with factor 0.2 initial learning rate was determined as 10^{-3}, a learning rate of 4.10^{-5} was reached at the end of the training.

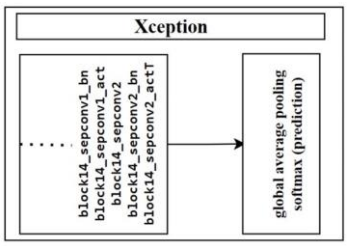

Figure 7. Layers added to the Xception model

The model that gave the most accurate result on the test data set among the checkpoint models recorded as a result of the training of all models was used for comparison.

4. COMPARISON OF TRAINING RESULTS AND MODELS

The models, which were trained by fine-tuning with the transfer learning method, were evaluated on the test data set as a result of the training. Precision, recall, accuracy and F1 scores obtained by the models on the test data were calculated and compared.

As a result of the evaluation made on the test data set for the VGG16 model, it was obersed that the best result was obtained with the model that gave the highest accuracy rate during the training, and this model was preferred for comparison. The complexity matrix showing the success of the VGG16 model in classifying the test data set is given in Figure 8.

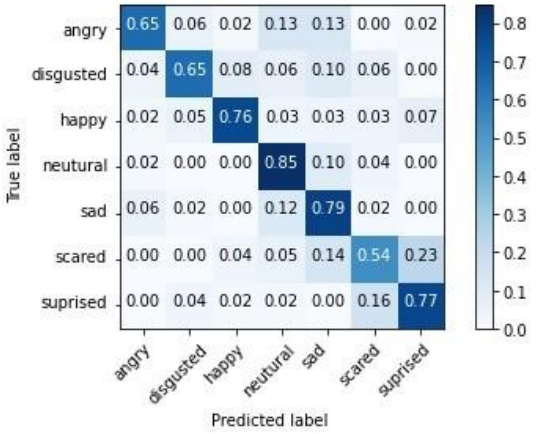

Figure 8. Confusion matrix of VGG16

When the complexity matrix given in Figure 8 is examined, it is seen that the VGG16 architecture classifies the neutral class as the most successful and the fear class as the most unsuccessful. Using the predictions made by the model on the test data set, accuracy, precision, sensitivity and F1 score values were calculated to evaluate the architecture and are given in Table 2.

Table 2. Accuracy, precision, sensitivity and F1 score results of VGG16 architecture

	Accuracy (Test Data)	Precision	Recall	F1 Score
VGG16	0.712	0.723	0.712	0.711

Although the trained VGG16 architecture achieved the highest accuracy rate of 72% on the accuracy dataset, this rate was found to be 71% on the test dataset. Again, on the test data set, 0.723 precision, 0.712 sensitivity and 0.711 F1 score values were obtained.

For the ResNet50 model, the best result on the test data set was obtained with the model that gave the highest accuracy rate recorded with the checkpoint. Therefore, the comparison was made over this model. The complexity matrix showing the success of the ResNet50 model in classifying the test data set is given in Figure 9.

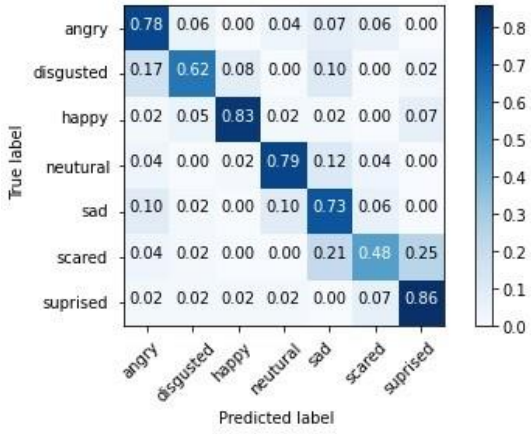

Figure 9. Confusion matrix of ResNet50

When the complexity matrix of the ResNet50 model is examined, it is seen that the model classifies the happy class as the most successful and the fear class as the most unsuccessful. Using the predictions made by the model on the test dataset, accuracy, precision, sensitivity and F1 score values were calculated to evaluate the architecture. The values calculated for the ResNet50 architecture are given in Table 3.

Table 3. Accuracy, precision, sensitivity and F1 score results of ResNet50 architecture

	Accuracy (Test)	**Precision**	**Recall**	**F1 Score**
ResNet50	0.728	0.735	0.726	0.724

The highest accuracy rate of 72% was obtained on the trained ResNet50 architecture accuracy dataset. Again, an accuracy rate of 72% was obtained on the test data set. In addition, 0.735 precision, 0.726 sensitivity and 0.724 F1 score values were obtained on the test data set.

As a result of the evaluation made on the test data set with the DenseNet121 model, it was seen that the best result was obtained with the model obtained as a result of the whole training, and this model was used for comparison. The test data set complexity matrix of the DenseNet121 model is shown in Figure 10.

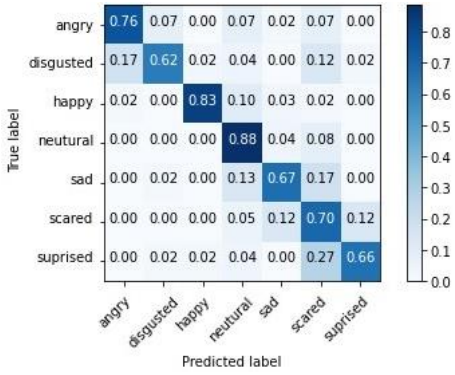

Figure 10. Confusion matrix of DenseNet121

When the complexity matrix is examined, it is seen that the DenseNet121 architecture classifies the neutral class as the most successful and the disgust class as the most unsuccessful. Using the predictions made by the model on the test dataset, accuracy, precision, sensitivity and F1 score values were calculated to evaluate the architecture. The values calculated for the DenseNet121 architecture are given in Table 4.

Table 4. Accuracy, precision, sensitivity and F1 score results of DenseNet121 architecture

	Accuracy (Test)	**Precision**	**Recall**	**F1 Score**
DenseNet121	0.734	0.762	0.732	0.738

Although the trained DenseNet121 architecture achieved the highest accuracy rate of 74% on the accuracy dataset, this rate was found to be 73% on the test dataset. Again, on the test data set, 0.762 precision, 0.732 sensitivity and 0.738 F1 score values were obtained.

For the InceptionV3 model, the best result on the test data set was obtained with the model that gave the highest accuracy rate recorded with checkpoint, and this model was used for comparison. The test data set complexity matrix of the InceptionV3 model is shown in Figure 11.

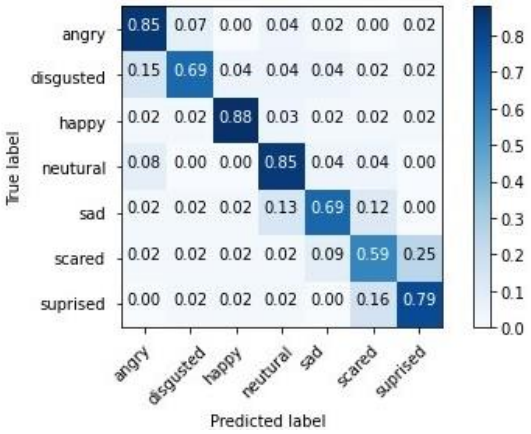

Figure 11. Confusion matrix of InceptionV3

When the complexity matrix shown in Figure 10 is examined, it is seen that the InceptionV3 architecture classifies the happiest class as the most successful and the fear class as the most unsuccessful. Using the predictions made by the model on the test data set, accuracy, precision, sensitivity and F1 score values were calculated to evaluate the architecture and are presented in Table 5.

Table 5. Accuracy, precision, sensitivity and F1 score results of InceptionV3 architecture

	Accuracy (Test)	**Precision**	**Recall**	**F1 Score**
InceptionV3	0.763	0.764	0.761	0.760

Although the highest accuracy rate of 74% was obtained on the accuracy dataset with the trained InceptionV3 architecture, this rate was found to be 76% on the test dataset. Again, on the test data set, 0.764 precision, 0.761 sensitivity and 0.760 F1 score values were obtained.
As a result of the evaluation made on the test data set for the InceptionResNetV2 model, it was seen that the best result was obtained with the model obtained as a result of the whole training, and this model was used for

comparison. The test data set complexity matrix of the InceptionResNetV2 model is shown in Figure 12.

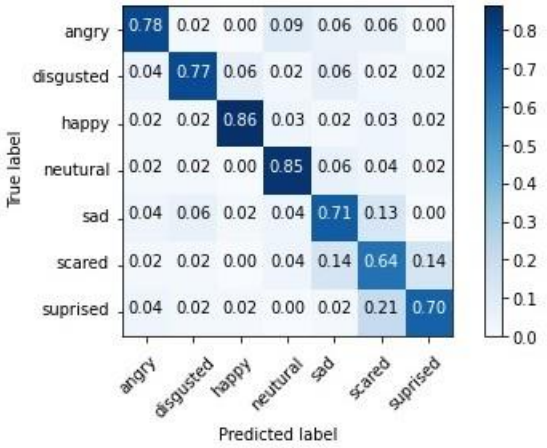

Figure 12. Confusion matrix of InceptionResNet50

When the complexity matrix given in InceptionResNetV2 is examined, it is seen that the InceptionResNetV2 architecture classifies the happiest class as the most successful and the fear class as the most unsuccessful. Using the predictions made by the model on the test dataset, accuracy, precision, sensitivity and F1 score values were calculated to evaluate the architecture. The calculated values for the InceptionResNetV2 architecture are given in Table 6.

Table 6. Accuracy, precision, sensitivity and F1 score results of InceptionResNetV2 architecture

	Accuracy (Test)	**Precision**	**Recall**	**F1 Score**
InceptionResNetV2	0.757	0.764	0.758	0.760

Although the highest accuracy rate of 71% was obtained on the accuracy dataset with the trained InceptionResNetV2 architecture, this rate was obtained as 75% on the test dataset. Again, on the test data set, 0.764 precision, 0.758 sensitivity and 0.760 F1 score values were obtained.

As a result of the evaluation made on the test data set for the Xception model, the best result was obtained with the checkpoint model with the lowest validation loss during the training. Confusion matrix obtained from the test data set using this model is given in Figure 13.

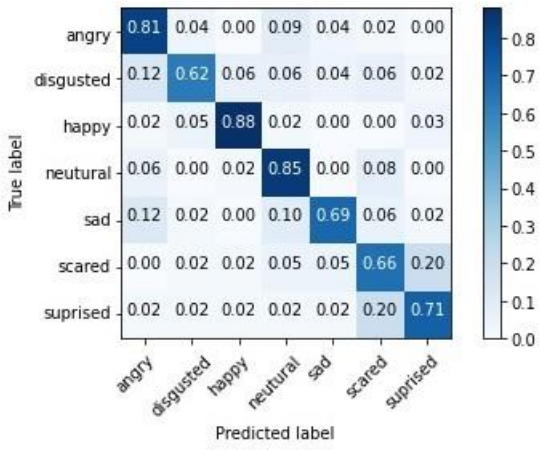

Figure 13. Confusion matrix of Xception

When the complexity matrix is examined, it is seen that the Xception architecture classifies the happy class as the most successful and the disgust class as the most unsuccessful. Using the predictions made by the model on the test dataset, accuracy, precision, sensitivity and F1 score values were calculated to evaluate the architecture. The values calculated for the Xception architecture are given in Table 7.

Table 7. Xception architecture accuracy, precision, sensitivity and F1 score results

	Accuracy (Test)	**Precision**	**Recall**	**F1 Score**
Xception	0.75	0.755	0.747	0.747

The highest accuracy rate of 73% was obtained on the trained Xception architecture accuracy dataset. In the test data set, the accuracy rate was found to be 75%. Again, on the test data set, 0.755 precision, 0.747 sensitivity and 0.747 F1 score values were obtained. The evaluation results of all models are given in Table 8.

Table 8. Evaluation results of models

Evaluation criteria	**Models**	**Value**
Recall	VGG16	0.712
	ResNet50	0.726
	DenseNet121	0.732
	InceptionV3	**0.761**
	InceptionResNetV2	0.758
	Xception	0.747
Precision	VGG16	0.723
	ResNet50	0.735
	DenseNet121	0.762
	InceptionV3	**0.764**
	InceptionResNetV2	**0.764**
	Xception	0.755
Accuracy	VGG16	71%
	ResNet50	73%
	DenseNet121	73.40%
	InceptionV3	**76.30%**
	InceptionResNetV2	75.70%
	Xception	75%
F1 score	VGG16	0.711
	ResNet50	0.724
	DenseNet121	0.738
	InceptionV3	**0.760**
	InceptionResNetV2	**0.760**
	Xception	0.747

When the precision, recall, acuracy and F1 score values calculated from the test data are examined; It is seen that the highest recall and accuracy values are obtained with the InceptionV3 model. In F1 score and precision values, it is seen that InceptionV3 and InceptionResNetV2 models are very close to each other or give the same results.

5. RESULTS AND DISCUSSION

In the study, it was tried to perform emotion analysis by predicting children's facial expressions with artificial intelligence models. In the study carried out, it was tried to compare the models by using different artificial intelligence models in describing facial expressions.

Although there are many studies on facial expression recognition in the literature, expression recognition studies on children's faces are very limited. For this purpose, it was aimed to detect expression on children's faces in this study. For this purpose, images of search results made using keywords on search engines were collected. Face detection was performed on the collected images, and the images were divided into seven classes according to the expressions they displayed: angry, disgusted, happy, neutral, sad, afraid and surprised.

Using 6 different deep learning models and transfer learning on the prepared data set, the model that gave the best results on the data set was determined. All the models used were able to achieve over 70% accuracy in the test dataset. Inception-based models seem to give very successful results.

When the F1 score values of the models given in Figure 14 are examined according to the classes, it is seen that all models are quite successful in detecting the happy class. In addition, it is seen that there is a decrease in all models in the detection of scared, surprised and sad classes. It is thought that this may be due to the fact that the pictures in the fear and surprise classes are largely similar to each other.

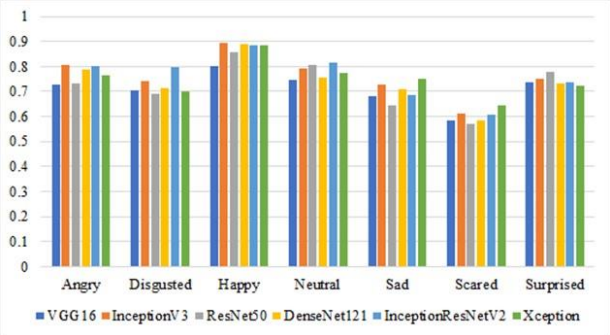

Figure 14. Class-based F1 scores of models

Since facial expression recognition is a very difficult and complex task, it is very important to have a large number of data in order to increase the success of the models. For this reason, datasets containing limited number of child facial expressions should be increased. Increasing the number of data of scared, sad and surprised facial expressions, which have low identification rates, will greatly benefit the success of the models.

In future studies, it is planned to expand the data set and detect real-time images. In addition, it is thought that training the models by using adult and child data sets together will increase the detection success of the models.

Acknowledgement

We would like to thank the open access websites for making the images available on different websites and allowing us to use them in the study.

Authors Contributions

The study was produced from the master thesis made by İrem SAYIN under the supervision of Bekir AKSOY

REFERENCES

[1] Jack R.E., Schyns P.G. The Human Face as a Dynamic Tool for Social Communication. Curr Biol. 2015; 25:R621–R634. https://doi.org/10.1016/j.cub.2015.05.052.
[2] DeVito Jospeh A. Human Communication. Boston: Pearson; 2002.
[3] Howard A., Zhang C., Horvitz E. Addressing bias in machine learning algorithms: A pilot study on emotion recognition for intelligent systems. 2017 IEEE Workshop on Advanced Robotics and its Social Impacts, ARSO 2017. Austin, TX, USA; 2017. https://doi.org/10.1109/ARSO.2017.8025197.
[4] Guo G., Guo R., Li X. Facial expression recognition influenced by human aging. IEEE Trans. Affect. Comput. 2013; 4: 291–298. https://doi.org/10.1109/T-AFFC.2013.13.
[5] Houstis O., Kiliaridis S. Gender and age differences in facial expressions. Eur. J. Orthod. 2009; 31: 459–466. https://doi.org/10.1093/ejo/cjp019.
[6] Brandao M., Age and gender bias in pedestrian detection algorithms. arXiv Prepr. arXiv:1906.10490, 2019.
[7] Egger H.L., Pine D.S., Nelson E., Leibenluft E., Ernst M., Towbin, K.E., et al. The NIMH Child Emotional Faces Picture Set (NIMH-ChEFS): a new set of children's facial emotion stimuli. Int. J. Methods Psychiatr. Res. 2011; 20: 145–156. https://doi.org/10.1002/mpr.343.
[8] Lobue V., Thrasher C., Kret M.E. The Child Affective Facial Expression (CAFE) set: validity and reliability from untrained adults. Front. Psychol. 2015; 5: 1532. https://doi.org/10.3389/fpsyg.2014.01532.
[9] Ekman P., Friesen W. V., EllsWorth P. Emotion in the Human Face. 1st ed. Pergamon Press; 1972. https://doi.org/10.1016/C2013-0-02458-9.
[10] Rao A., Ajri S., Guragol A., Suresh R., Tripathi S. Emotion Recognition from Facial Expressions in Children and Adults Using Deep Neural Network. Int. J. Intell. Syst. 2020; 43–51. https://doi.org/10.1007/978-981-15-3914-5_4.

[11] Leo M., Del Coco M., Carcagnì P., Distante C., Bernava M., Pioggia G., et al. Automatic Emotion Recognition in Robot-Children Interaction for ASD Treatment. IEEE International Conference on Computer Vision, ICCV 2015. Santiago, Chile: 2015.p 537–545. https://doi.org/10.1109/ICCVW.2015.76.

[12] Nagpal, S., Singh, M., Vatsa, M., Singh, R., Noore, A. Expression classification in children using mean supervised deep Boltzmann Machine. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, CVPR 2019. California: 2019.

[13] Witherow, M. A., Samad, M. D.,Iftekharuddin, K. M. Transfer learning approach to multiclass classification of child facial expressions. SPIE Optical Engineering + Applications. San Diego, California, United States: 2019. p. 1113911

[14] Lopez-Rincon A. Emotion recognition using facial expressions in children using the NAO robot. 2019 International Conference on Electronics, Communications and Computers , CONIELECOMP 2019. Cholula, Mexico:IEEE; 2019.p.146-153. 10.1109/CONIELECOMP.2019.8673111