Emotion detection using facial expression recognition based on VGG16 network

RENJU RENJITH[1*], Dr JOBY PP[2]

[1]M techStudent, [2] Professor

Computer Science and Engineering,

St Joseph's college of Engineering and Technology, Palai, Kottayam, kerala, India.

***Abstract*:** Facial Expression Recognition is an innovation which utilizes biometric markers to recognize feelings in human faces. All the more unequivocally, this innovation is a supposition investigation device and can naturally distinguish the six fundamental or general articulations.. Facial Expression is one of the significant nonverbal channels through which human feeling state is conveyed, it includes the examination and acknowledgment of facial highlights. Facial Expression Recognition is classified as social biometrics and furthermore relevant in the field of computer vision . The traditional machine learning algorithms produce high accuracies for similar tasks, they lack to detect emotions of faces, or in different poses or environmental conditions. In this paper, a convolutional neural network architecture, VGG16 network , is proposed to address the FER problem by using a dataset. The proposed architecture is used to detect the emotion of the students in the online class using the facial expression .According to the experimental result, the accuracy of architecture was calculated to be as high 96% on a de-facto standard dataset, namely CK+. The experimental result confirmed that proposed neural network architecture is fast enough to be integrated into real -time FER application .

Index terms: Artificial intelligence, artificial neural networks, Convolutional Neural Network,Facial expression Recognition, Local Binary Pattern

I.INTRODUCTION

Emotion detection assumes a significant part in numerous spaces like clever security , mechanical technology fabricating, clinical brain research, media, and automotive security .Look acknowledgment (FER), which is a significant examination space of Human Machine Interaction (HMI), is the assignment of recognizing feelings by dissecting facial expression that assume a critical part in social collaboration [23] and pass on significant and clear data about the feelings of individuals [1]. As a characteristic outcome of that, different computer vision frameworks dependent on AI calculations have proposed FER where they were prepared utilizing explained face datasets. FER applications take the photographs of subjects as the information and produce the distinguished feelings through different examinations as the yield. The objective feelings fluctuate through the proposed approach which could be happiness, sadness, surprise, anger, disgust, fear, contempt, and neutral. The overall design of FER approaches contains three stages, to be specific, (1) preprocessing, (2) extraction, and (3) classification stage. In the first stage, the preprocessing stage, the nature of the information pictures is upgraded and the repetitive data is taken out. In the extraction phase, the preprocessed input information are changed into the best delegate highlights to lead a touchy and adaptable grouping method that could play out the correct forecasts as far as feelings. In the last stage mapping of input data to the target emotions takes place by virtue of the utilized classification algorithm. Characterization is accomplished by some uncommon grouping calculations, for example, Support Vector Machine (SVM), AdaBoost calculation, Artificial Neural Network (ANN) to specify a group. As the element extraction is significant for acceptable framework execution so additionally is the arrangement calculation carries out. Deep learning networks are more capable for extracting features from the training data and it provide more accuracy in emotion detection.For the proposed architecture ,CK+ dataset was used to examine the result of the emotion detection. Facial expression Recognition take the input image and classify the facial expression into six, which is happiness, sadness, surprise, anger, disgust and fear.

The main steps of the proposed system is as follows:

• The VGG16 architecture, which was specifically designed to detect the facial expression for a given photo in the tolerable limit to latency for real-time applications, was proposed whose accuracy was calculated to be as high as 95% on CK+ dataset

• The proposed architecture was implemented using open-source software (e.g. various Python libraries, OpenCV, Keras , Tensorflow, tkinter etc.), which makes the model more flexible to the proposed system.

• The VGG16 architecture model is build and train the model with CK+ dataset ,then create a dataset using image dataset generator for emotion detection in real time. The dataset which is created by us is trained by the VGG16 model and classify the real image or photo into 6 facial expression.

The rest of the article is structured as follows: Section 2describes the literature surveys . Section 3 describes the proposed VGG16 architecture for detect the emotion using facial expression in real time. Section 4 presents the experimental result and discussion. Finally, Section 5 concludes the paper with future directions.

II.RELATED WORKS

LBP is ideal for texture analysis, the purpose at the back of LBP operator is that photograph texture may be represented through the nearby spatial and the grey scale comparison that are seemed as complementary measures [2]. The main advantage while using Local Binary Pattern is its ,simple way of computation, high tolerance for low image resolution, discriminating prowess and high accuracy in varying illumination. Shahreen et al. [3] claimed that the use of LBP function on SVM classifier for seven expressions in JAFFE database progressed the machine overall performance through 25%.The speculation in the back of this system is that there are subsets of face segments which can be extra articulated in appearance investigation, getting the measurements courting of those segments ought to safely deal with the issue vectors [4].Geometric characteristic aren't prompted via way of means of lights condition. Liu et al. [6] proposed a calculation that fuse LBP and HOG features extricated from CK+ and JAFFE data set and decreased the removed features dimensionality with PCA, after permutated the combination on a few classifiers, he tracked down that the melded includes on softmax classifier created 98.3 percentage on CK+ and 90 percentage on JAFFE information base. The outcome is a proof that a legitimate half and half features could fundamentally improve framework execution. Majumder et al. [18] proposed a Facial Expression Recognition framework has 4 main stages as follows: (1) feature extraction by using geometric features (2) Feature extraction by LBP (3) Fusion of both features using encoders and the last stage (4) classification by SVM. By the properly defined labels SVM has proved to achieve success in recognizing facial expression[13]. The main advantages of SVM is high performance in phrases of accuracy and flexibility of the data. The main challenge faced by the SVM accuracy via way of means of the presence of a lot coarse noise withinside the label [13]. Convolutional neural networks, which might be a category of deep feed forward neural networks, have yielded extremely good performancesn lots of troubles in computer vision withinside the beyond years Advantage of the use of a deep neural community in preference to a conventional system studying set of rules is that once deep neural networks are skilled on huge datasets, they extract generalized capabilities, which can be carried out on datasets that the network has now no longer been trained in [23]

table 1: Summary of the FER Methods

AUTHOR	METHOD (FEATURE and CLASSIFIER)	PERFORMANCE	LIMITATION
Feng et al [20]	Local Binary Pattern + Linear Programming	93.8%	Posed dataset for system training will mislead the system in real time environment
G. Sandbach et al. [21]	3D motion based feature + Hidden Markov Model	81.93%	Being a generative model it experience Local optimal Solution trap
Y. Liu et al [17].	Hybrid (LBP&HOG) +Softmax	96.3%	Research is still finding best combinations of features.
D. Huang et al [18].	Local Binary Pattern + AdaBoost	84.6%	Loss of relevant information is possible
Abdullah Talha et al [2]	Convolutional Neural Network	97.06%	Tendency to overfit
I. Kotsia and I. Pitas [15]	Geometric Deformation Features +Support Vector Machine	90.1%	Generalization Problem.

III. MATERIAL AND METHOD

In this section, described detailed about the dataset which is used to evaluate the proposed system and the detailed description about the proposed architecture of the system for emotion detection using facial expression in realtime.

3.1Dataset

The Extended Cohn-Kanade Dataset (CK+) is a standard FER benchmark dataset that consists of 210 subjects at the ages of 18 to 50 years. The dataset contains a great variety of photos of both genders with different ages having different backgrounds. The sizes of the photos in this dataset are 640× 480 and 640 × 490 pixels. The dataset contains both colored and grayscale photos and consists of 327 sequences from 123 subjects. Each sequence is categorized into one of the seven facial emotions which are (1) anger, (2) disgust, (3) fear, (4) happiness, (5) sadness, (6) surprissse, and (7) contempt. The number of frames in a sequence varies from 10 to 60 as each subject performs the target facial expression transitions from the neutral phase.

3.2proposed architecture

VGG16 is a convolution neural net (CNN) architecture which was used to win ILSVR(Imagenet) competition in 2014. It is considered to be one of the excellent vision model architecture till date. Most unique thing about VGG16 is that instead of having a large number of hyper-parameter they focused on having convolution layers of 3x3 filter with a stride 1 and always used same

padding and maxpool layer of 2x2 filter of stride 2. It follows this arrangement of convolution and max pool layers consistently throughout the whole architecture. In the end it has 2 FC(fully connected layers) followed by a softmax for output. The 16 in VGG16 refers to it has 16 layers that have weights. This network is a pretty large network and it has about 138 million (approx) parameters.

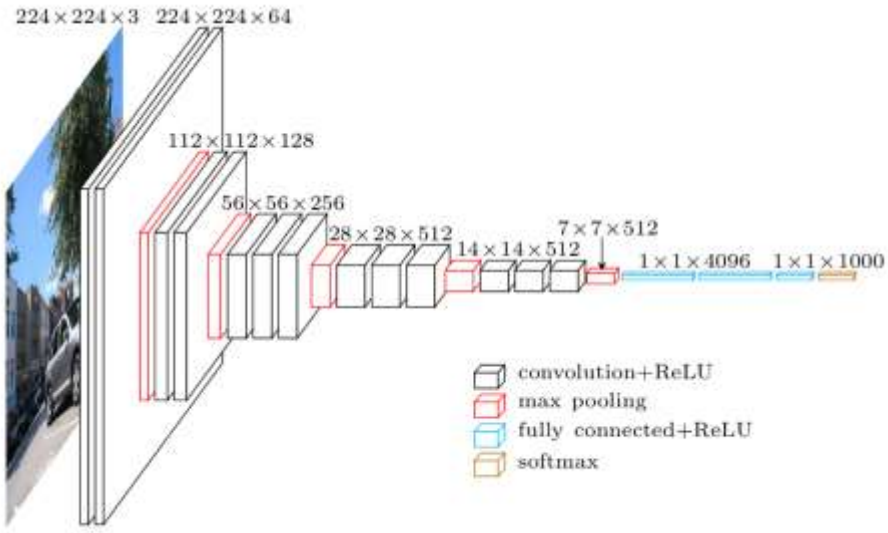

fig 2: VGG16 architecture

This network is a pretty large network and it has about 138 million (approx) parameters.Importing necessary libraries for build the VGG16 model.Train the model using the dataset .it take almost 5-6 hours for the training process.I first importing all the libraries which i will need to implement VGG16. I will be using Sequential method as I am creating a sequential model. Sequential model means that all the layers of the model will be arranged in sequence. Here I have imported ImageDataGenerator from keras.preprocessing. The objective of ImageDataGenerator is to import data with labels easily into the model. It is a very useful class as it has many function to rescale, rotate, zoom, flip etc. The most useful thing about this class is that it doesn't affect the data stored on the disk. This class alters the data on the go while passing it to the model.creating and object of ImageDataGenerator for both training and testing data and passing the folder which has train data to the object trdata and similarly passing the folder which has test data to the object tsdata.

3.2.1Emotion detection real time

Test the model that we build for emotion detection in real-time using OpenCV and webcam. To do so we will write a python script. We will use the spyder in our local system to make use of a webcam.After importing all the required libraries we will load the model weights that we saved earlier after training.After importing the model weights we have imported a haar cascade file that is designed by open cv to detect the frontal face.After importing the haar cascade file we will have written a code to detect faces and classify the desired emotions. We have assigned the labels that will be different emotions like angry, happy, sad, surprise, neutral. As soon as you run the code a new window will pop up and your webcam will turn on. It will then detect the face of the person, draw a bounding box over the detected person, and then convert the RGB image into grayscale classify it in real-time

IV.EXPERIMENTAL RESULTS

The proposed deep convolutional neural network architecture was evaluated by CK+ dataset in order Before training the proposed deep convolutional neural network, all the images in the CK+ dataset were cropped in order to remove the redundant parts of images such as background thanks to the OpenCV face detection module, namely, opencv python. Each sequence starts with a photo of the subject with a neutral facial expression and ends with a photo of the subject with the target facial expression. The proposed deep convolutional neural network was trained until the point the loss of training was saturated, which occurred around the epoch 200. Then, the training was stopped, and the testing phase was started. The calculated losses for both the validation and training phases. After building the model using VGG16 architecture and trained the dataset and test the dataset and create a .yml file.Creating the dataset using the dataset generator,take the screenshot of the photos using webcam and crop each of the face from the photo and detect the emotion of the each face using the facial expression.

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 224, 224, 64)	1792
conv2d_2 (Conv2D)	(None, 224, 224, 64)	36928
max_pooling2d_1 (MaxPooling2	(None, 112, 112, 64)	0
conv2d_3 (Conv2D)	(None, 112, 112, 128)	73856
conv2d_4 (Conv2D)	(None, 112, 112, 128)	147584
max_pooling2d_2 (MaxPooling2	(None, 56, 56, 128)	0
conv2d_5 (Conv2D)	(None, 56, 56, 256)	295168
conv2d_6 (Conv2D)	(None, 56, 56, 256)	590080
conv2d_7 (Conv2D)	(None, 56, 56, 256)	590080
max_pooling2d_3 (MaxPooling2	(None, 28, 28, 256)	0
conv2d_8 (Conv2D)	(None, 28, 28, 512)	1180160
conv2d_9 (Conv2D)	(None, 28, 28, 512)	2359808
conv2d_10 (Conv2D)	(None, 28, 28, 512)	2359808
max_pooling2d_4 (MaxPooling2	(None, 14, 14, 512)	0
conv2d_11 (Conv2D)	(None, 14, 14, 512)	2359808
conv2d_12 (Conv2D)	(None, 14, 14, 512)	2359808
conv2d_13 (Conv2D)	(None, 14, 14, 512)	2359808
max_pooling2d_5 (MaxPooling2	(None, 7, 7, 512)	0
flatten_1 (Flatten)	(None, 25088)	0
dense_1 (Dense)	(None, 4096)	102764544
dropout_1 (Dropout)	(None, 4096)	0
dense_2 (Dense)	(None, 4096)	16781312
dropout_2 (Dropout)	(None, 4096)	0
dense_3 (Dense)	(None, 2)	8194

Total params: 134,268,738
Trainable params: 134,268,738
Non-trainable params: 0

fig 3: VGG16 model build

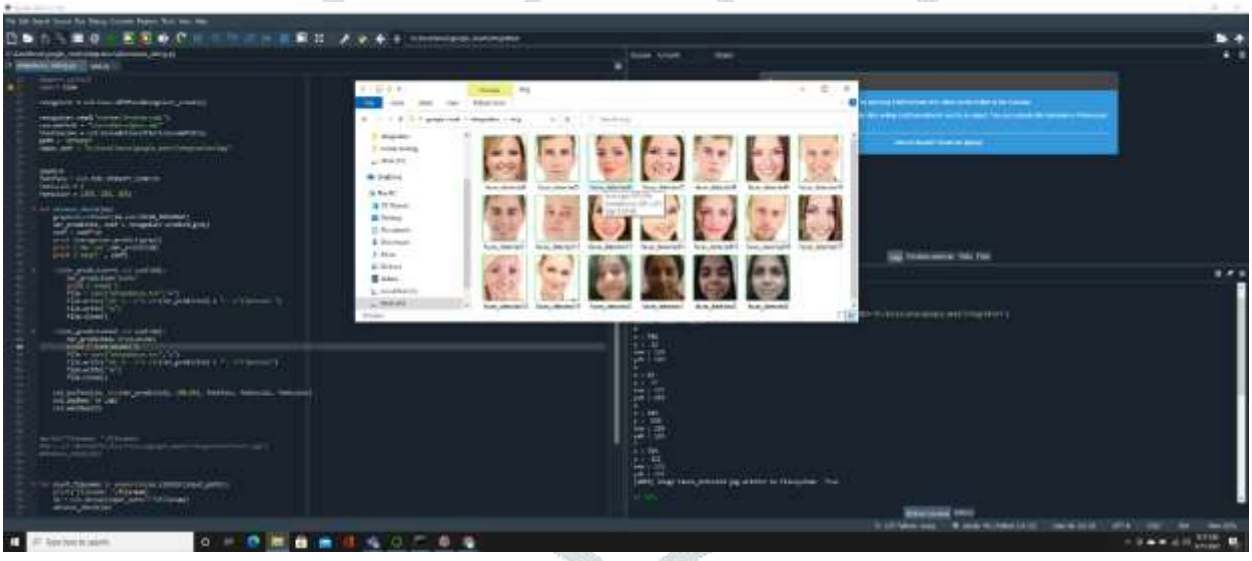

fig 4: output of training the model

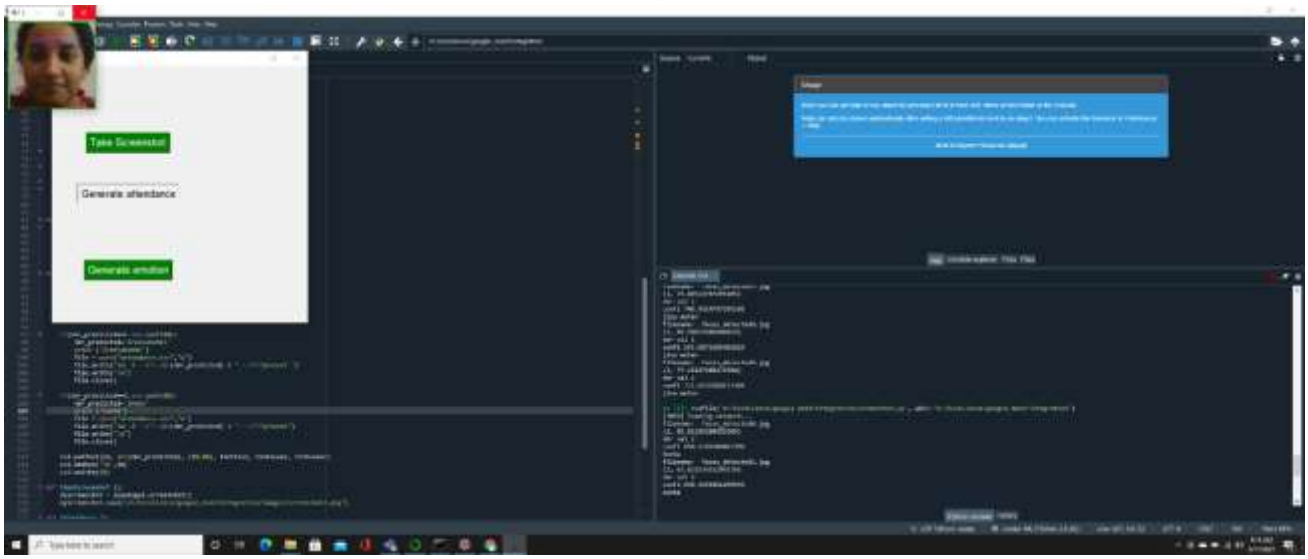

fig5: output of the proposed architecture

V.CONCLUSION

The goal of face emotion recognition (FER) is figuring out feelings of a human. The emotion may be captured both from face or from verbal communication. Psychological traits including heartbeat and blood Pressure, speech, hand gestures, frame movements, Facial expressions pick out emotions of a person. Facial emotion reputation is one of the beneficial tasks and may be used as a base for lots real-time packages may be used as part of many thrilling and beneficial packages like Monitoring security, treating sufferers in scientific field, advertising research, E-learning etc;. We human beings can easily pick out the emotion of different human beings with none effort. Automatic detection of emotion of a human face is vital because of its use in real-time applications In future examinations, more effective hand crafted features can be incorporated in to the existing methods of facial expression recognition .Also, cross data set preparing network boundaries can be utilized to improve speculation capacity.

References

[1] O. Ekundayo and S. Viriri, “Facial expression recognition: A review of methods, performances and limitations,” in Proc. Conf. Inf. Commun. Technol. Soc. (ICTAS), Mar. 2019, pp. 1–6, doi: 10.1109/ICTAS. 2019.8703619.

[2] T. Ojala, M. Pietik¨ainen, and T. M¨aenp¨a¨a, “Gray Scale and Rotation Invariant Texture Classification with Local Binary Patterns,” Springer, Berlin, Heidelberg, 2000, pp. 404–420.

[3] S. Kasim, R. Hassan, N. H. Zaini, A. S. Ahmad, and A. A. Ramli, “A Study on Facial Expression Recognition Using Local Binary Pattern,” Int. J. Adv. Sci. Eng. Inf. Technol., vol. 7, no. 5, pp. 1621– 1626, 2017

[4] B. Martinez, M. F. Valstar, B. Jiang, and M. Pantic, Automatic Analysis of Facial Actions: A Survey,” IEEE Trans. Affect. Comput., vol. 3045, no. c, 2017

[5] D. Ghimire and J. Lee, “Geometric Feature-Based Facial Expression Recognition in Image Sequences Using Multi-Class AdaBoost and Support Vector Machines,” Sensor, vol. 13, pp. 7714–7734, 2013

[6] Y. Liu, Y. Li, X. Ma, and R. Song, “Facial Expression Recognition with Fusion Features,” Sensor, vol. 712, no. 17, pp. 1–18, 2017.

[7] A. Majumder, L. Behera, and V. K. Subramanian, “Automatic facial expression recognition system using deep network-based data fusion,” IEEE Trans. Cybern., vol. 48, no. 1, pp. 103–114, Jan. 2018, doi: 10.1109/ TCYB.2016.2625419

[8] M. Dahmane and J. Meunier, “Emotion Recognition using Dynamic Grid-based HoG Features,” IEEE Xplore, pp. 884–888, 2011.

[9] P. Kumar, \A Real-time Robust Facial Expression Recognition System using Features," Int. Conf. Comput. Anal. Secur. Trends, pp. 289{293, 2016

[10] Y. L. Tian, T. Kanade, and J. F. Cohn, \Facial expression analysis," Handb. Face Recognit., pp. 247{275, 2005.

[11] G. Fung and O. L. Mangasarian, \Proximal support vector machine classi_ers," inProceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '01, 2001, pp. 77{86.

[12] G. M. Fung and O. L. Mangasarian, \Multicategory Proximal Support Vector Machine Classi_ers," Mach. Learn., vol. 59, no. 1{ 2, pp. 77{97, 2005.

[13] S. W. Chew, S. Lucey, P. Lucey, S. Sridharan, and J. . Conn, \Improved FacialExpression Recognition via Uni-Hyperplane Classi_cation," IEEE Comput. Soc.Conf. Comput. Vis. Pattern Recognit., pp. 2554{2561, 2012.

[14] H. H. Zhang, Y. Liu, Y. Wu, and J. Zhu, \Variable selection for the multicategorySVM via adaptive sup-norm regularization," Electron. J. Stat., vol. 2, pp. 149{167,2008.

[15] Z. Wang and X. Xue, \Multi-Class Support Vector Machine," in Support VectorMachines Applications, Cham: Springer International Publishing, 2014, pp. 23{48.

[16] T. Zhang, W. Zheng, Z. Cui, Y. Zong, J. Yan, and K. Yan, \A Deep Neural Network-Driven Feature Learning Method for Multi-view Facial Expression Recognition,"IEEE Trans. Multimed., vol. 18, no. 12, pp. 2528{2536, Dec. 2016

[17] R. Breuer and R. Kimmel, \A Deep Learning Perspective on the Origin of FacialExpressions," arXiv, vol. 5, no. 5, pp. 1{16, 2017.

[18] D. Huang, C. Shan, M. Ardebilian, Y. Wang, and L. Chen, \Local Binary Patternsand Its Application to Facial Image Analysis: A Survey," IEEE Trans. Syst. Man,Cybern. Part C, vol. 41, no. 6, pp. 765{781, 2011

[19] T. Ojala, M. Pietik • ainen, and T. M • aenp • a • a, \Gray Scale and Rotation InvariantTexture Classi_cation with Local Binary Patterns," Springer, Berlin, Heidelberg,2000, pp. 404{420.

[20] X. Feng, M. Pietik • ainen, and A. Hadid, \Facial Expression Recognition Based onLocal Binary Patterns 1," Pattern Recognit. Image Anal., vol. 17, no. 4, pp. 592{598,2007.

[21] G. Sandbach, S. Zafeiriou, M. Pantic, and D. Rueckert, \A Dynamic Approachto the Recognitionof 3D Facial Expressions and Their Temporal Models," FaceGesture, pp. 406{413, 2011.

[22] K. Sikka, A. Dhall, and M. Bartlett, \Exemplar Hidden Markov Models for classication of facial expressions in videos," Comput. Vis. Pattern Recognit. Work.(CVPRW), 2015 IEEE Conf., pp. 18{ 25, 2015

[23] Mollahosseini, D. Chan, and M. H. Mahoor, \Going deeper in facial expressionrecognition using deep neural networks," in Proc. IEEE Winter Conf. Appl. Comput.Vis. (WACV), Mar. 2016, pp. 1{10, doi: 10.1109/WACV.2016.7477450.

[24] M. F. Valstar, M. Pantic, and S. Member, \Fully Automatic Recognition of the TemporalPhases of Facial Actions," IEEE Trans. Syst. Man, Cybern. Part B Cybern.,vol. 42, no. 1, pp. 28{ 43, 2012.