An Integrated Real-Time Deep Learning and Belief Rule Base Intelligent System to Assess Facial Expression Under Uncertainty

Tawsin Uddin Ahmed[1], Mohammad Newaj Jamil[2], Mohammad Shahadat Hossain[3],
Karl Andersson[4], Mohammed Sazzad Hossain[5]
[1 2 3] Department of Computer Science and Engineering, University of Chittagong, Chittagong, Bangladesh
[4]Department of Computer Science, Electrical and Space Engineering, Luleå University of Technology, Skellefteå, Sweden
[5]Department of Computer Science and Engineering, University of Liberal Arts, Bangladesh
Email: [1]tawsin.uddin@gmail.com, [2]hridoyjamil10@gmail.com, [3]hossain_ms@cu.ac.bd,
[4]karl.andersson@ltu.se, [5]sazzad.hossain@ulab.edu.bd

***Abstract*—Nowadays, the recognition of facial expression draws significant attention in various domains. In view of this, a real-time facial expression recognition system has been developed using a Deep Learning approach, which can classify ten emotions, including angry, disgust, fear, happy, mockery, neutral, sad, surprise, think, and wink. In addition, an integrated expert system has also been developed by integrating Deep Learning with a Belief Rule Base to support the assessment of the overall mental state of a person over a period of time from video streaming data under uncertainty. In this research, data-driven and knowledge-driven approaches are integrated together to assess the mental state of an individual. Such a system could enable the identification of a suspect before committing any crime beforehand by the law enforcement agency. The performance of this integrated system is found reliable than existing methods of facial expression assessment.**

***Contribution*—The paper presents a noble method of computing the overall mental condition of a person by integrating CNN and BRBES under uncertainty.**

***Keywords*—Facial Expression Recognition, Deep Learning, Belief Rule Base, Integrated Framework, Uncertainty**

I. INTRODUCTION

The most frequent way that most of the people like to express their emotions through facial expressions. Basically, it refers to the gesture of face muscles to visualize distinguish states of a face which are plotted as the reflection of an individual's emotion. However, sometimes it is difficult for humans let alone machines to analyze facial expressions since variation in expressing emotion through facial expressions exists. In spite of that, it has become one of the most interesting topics for conducting research. A Convolutional Neural Network (CNN) is a deep learning algorithm that is developed by imitating the biological visual cortex. CNN employs its convolution layer part to make a system understanding the form of particular visual content. Actually, a visual object can be represented as its unique features or patterns. That is what the Convolution layer extracts from an object and passes to the regular neural network. CNN is now vastly using in most computer vision problems mostly in image classification [1]. CNN requires training on a dataset that makes a machine to produce accurate decisions even on unseen data. The arrival of CNN boosts the capability of a deep learning approach to a great extent that many computer vision problems get its way to a solution. Image recognition, speech recognition, text classification, semantic segmentation, analyzing satellite data are some of the applications of CNN [2][3]. A Belief Rule-Based Expert System (BRBES) is a knowledge-driven approach that is organized by a set of rules and delivers outputs based on the execution of those rules. Since the rules are the combination of antecedents and consequent, the antecedents of a rule gets triggered based on the input and generates the consequent part [4]. In this paper, the BRB part of the integrated CNN-BRB framework is responsible to assess the overall mental state of a person from the facial expression information provided by the Convolutional Neural Network. The remainder of this article is structured as follows: Section II covers related work on facial expression recognition, while Section III and Section V provides an overview of Convolutional Neural Network and Belief Rule-Based Expert System. Section VI discusses the Integration of the Convolutional Neural Network and Belief Rule-Based Expert System. Section VII describes the experiment process, and Section VIII presents an analysis of the results. Finally, Section IX concludes the paper.

II. RELATED WORKS

There has been some research performed on facial expression recognition: Deep sparse autoencoders [5] are developed to recognize facial expression which extracts geometric features from the data. Research work is tested on a popular dataset of facial expression cohn-kanade(CK+) database depicting seven expressions The proposed model is successful to achieve the accuracy of 95.79%. Lopes et al. [6] applied some image processing techniques on the facial expression image dataset to extract only the features related to facial expression. In the training phase, the model is trained based on the expression id along with the eye center location of the facial expression image to capture weight set for CNN.

The model is evaluated on three publicly available datasets which are CK+, JAFFE, and BU-3DEF and mentioned obtaining 96.76% accuracy in CK+ dataset. E. Owusu et al. [7] conducted a facial expression recognition process through some phases to improve recognition accuracy. Facial features are extracted by applying Gabor feature extraction techniques, which deliver a variety of facial deformation patterns. Furthermore, the formulation of the AdaBoost approach takes place to reduce the number of features that keep useful features only. According to [7], the system is trained and tested on two well-known datasets of FER named JAFFE and YALE. A. R. Rivera et al. [8] proposed a model that includes a face descriptor, which is named Local Direct Number Patter (LDN), to encode the neighborhood's structure with the help of its directional information. A compass mask is employed to generate edge responses in the neighborhood. Applying SVM for the classification task, the proposed model is performed on five facial expression datasets which are CK, CK+, JAFFE, MMI, and CMU-PIE.

III. Convolutional Neural Network

Convolutional Neural Network (CNN) is an Artificial Neural Network (ANN) that is so far been most popularly used for analyzing images. Although image analysis has been the most widespread use of CNN, it can also be used for other data analysis or classification problems as well [9]. The convolution layer is the layer that makes CNN distinguish from the regular ANN. In this research, CNN is used because it involves image datasets and CNN is well known to perform well for image classification tasks. However, CNN demands a large image dataset for better classification accuracy. Also, the accuracy of the classification task is proportional to the size of the image dataset. The system architecture is shown in Table I.

TABLE I: System Architecture

Model Content	Details
First and Second Convolution Layer	64 filters of size 3×3, ReLU, input size 48×48
Max Pooling Layer	Pooling Size 2×2, Strides of (2,2)
Third and Forth Convolution Layer	128 filters of size 3×3, ReLU
Max Pooling Layer	Pooling Size 2×2, Strides of (2,2)
Flatten Layer	Convert 2D matrix into 1D vector
First Hidden Layer	1024 nodes, ReLU
Dropout Layer	Deactivates 50% nodes randomly
Second Hidden Layer	1024 nodes, ReLU
Dropout Layer	Deactivates 50% nodes randomly
Output Layer	10 nodes for 10 classes, SoftMax
Optimization Function	AdaBound
Learning Rate	0.001
Callbacks	EarlyStopping, ReduceLROnPlateau, ModelCheckpoint, TensorBoard

According to Table I, the organization of CNN started with two convolution layers with 64 filters each having a size of 3×3 to extract facial features. Then, to fetch the maximum features from the feature maps MaxPooling layer with pool size 2×2 is included. Following the previous two layers, second and third convolution layers which are constructed by 128 filters of size 3×3 each. Like the previous pooling layer, this time also MaxPooling layer with pool size 2×2 is inserted. A flatten layer is embedded just after the last pooling layer. As a part of the fully connected layer, 2 hidden layers with 1024 nodes per layer are formulated. Hence, each hidden layer is followed by a dropout layer which randomly deactivates 50% of the active nodes so that the model can not be biased to the training data. The activation function that is applied in all convolution as well as hidden layers is ReLU. At last, CNN is concluded with an output layer which consists of 10 nodes for ten classes and softmax is used as activation function is this layer. To optimize the learning process AdaBound optimizer with .001 learning rate is applied [10]. Since overfitting is one of the concerned factors in CNN training, EarlyStopping is considered as one of the callbacks to stop its training when there is no improvement in learning.

IV. Data Collection and Preprocessing

The proposed Convolutional Neural Network is trained on a combined dataset collected from the different sources of standard datasets. The origins of this combined dataset are illustrated below:

- CK & CK+: Initial release which is known as CK contains 486 sequences from 97 posers started with neutral expression gradually proceeds to the peak expression. CK+ includes both posed and non-posed face images where the number of sequences swells up with 22% and the number of subjects with 27% [11].
- Fer2013 : This dataset is one of the most popular data in facial expression recognition. It is a collection of 35887 grayscale 48x48 images which is a combination of seven expressions - angry, disgust, fear, happy, sad, surprise, neutral [12].
- The MUG facial expression dataset: In this dataset 86 subjects including 35 women and 51 men performing six facial expressions which are anger, disgust, fear, happiness, sadness, surprise [13].
- KDEF & AKDEF: 4900 pictures of KDEF has been used more than 1500 research publications. It contains seven facial expressions - anger, disgust, fear, happiness, sadness, surprise [14].

In addition to seven regular facial expressions, three unique facial expressions, namely 'mockery', 'think', and 'wink' are added by downloading quality images from Shutterstock and adobe stock. Actually, in this research, a new dataset is going to be proposed for mockery, think and wink that might help other researchers in this domain.

In data preprocessing phase, the dataset of the proposed system has to undergo some steps. Initially, face detection is executed to extract the region of interest with the help of the Haar Cascade classifier [15]. The detected face region is cropped and converted to grayscale to reduce the color complexity of the image. In order to diminish low-intensity problem, the images of the dataset are normalized.

V. Belief Rule-Based Expert System

A Belief Rule-Based Expert System (BRBES) has two main components, namely a knowledge base and an inference engine [4]. Belief Rule Base (BRB) is utilized to create the initial knowledge base, while Evidential Reasoning (ER) works as an inference engine in a BRBES [16]. A belief rule is an extended form of traditional IF-THEN rule. There are two main parts in a belief rule, namely antecedent, and consequent. In a belief rule, each antecedent attribute takes referential values, while belief degrees are associated with the consequent attribute in a belief rule. Attribute weight, rule weight, and belief degrees, which are the knowledge representation parameters are used to capture uncertainty in data in a BRB [17]. The inference procedures contain various steps, including input transformation, rule activation, belief update, and rule aggregation, as shown in Fig. 1.

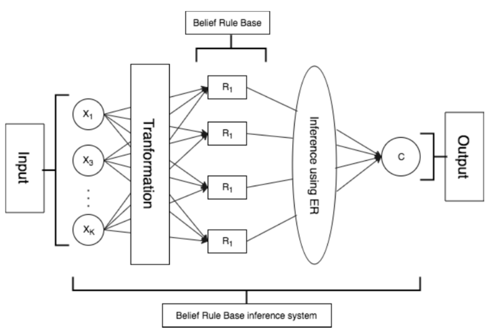

Fig. 1: Sequence of BRBES Inference Procedures

The input transformation distributes the input data over the referential values of the antecedent attributes of a belief rule, which is known as matching degrees [18]. After matching degrees are calculated, the belief rules are called packet antecedent, and they are considered active [19]. The matching degrees are used to calculate the activation weight of the rules. If an input data for any of the antecedent attribute is missing or ignored, the belief degree associated with each belief rule in the rule base should be updated [20]. Afterward, the aggregation of the rules is carried out by using either analytical or recursive evidential reasoning algorithm [21][22]. The final result is obtained by calculating the crisp value from the fuzzy output of rule aggregation procedure using the utility score associated with each referential value of the consequent attribute [23]. All these steps are performed by following the procedures mentioned in [24], [25].

VI. Integration of CNN and BRBES

Videos are nothing but a sequence of frames and for the assessment of facial expression, these frames are the inputs for our trained model. To extract image frames from a video, we applied an OpenCV function called cv2.VideoCapture(). Using the Haar Cascade classifier, the face portion of each frame is extracted and then convoluted by convolution layers of the integrated framework. Each filter of the convolution layer generates a feature map that is delivered to be pooled by the MaxPooling layer where maximum pixel value within a pooling window is picked up. Trained hidden layers with optimized weight for facial expression recognition perform matrix chain multiplication on the given inputs and forward their output to the output layer. In the output layer, with the help of Softmax activation function probabilities, which indicates the possibility that the frame belongs to that particular class, are distributed among the classes. Then the average probability of each class for total frames is calculated. These classes are utilized as the referential values of the antecedent attribute for the Belief Rule-Based Expert System (BRBES). In BRBES, “Person’s Expression” is considered as the antecedent attribute, and each class is considered as the corresponding referential values of the antecedent attribute. “Overall Mental State” is considered as the consequent attribute, its referential values are chosen as “Good”, “normal”, and “bad”, and its utility values are chosen as “1.0”, “0.5”, and “0.0”, respectively. The belief rule base used for BRBES is shown in Table II. Based on the belief rule base, the inference procedure is conducted, and the final output is obtained. The working process of the integration of the Convolutional Neural Network (CNN) and Belief Rule-Based Expert System (BRBES) is shown in Fig. 2.

VII. Experiments

Deep learning model training requires adequate GPU as the neural network is nothing but a huge number of matrix multiplication. Due to the limitation in terms of processing power in CPU, proposed CNN architecture is set for training on cloud server which is known as Google Colaboratory. Google Colab is specially built to let the machine learning researchers get out of the shortage problem of the processing unit by providing shared K80 GPU and Jupyter notebook environment as well. Despite collecting images from several standard datasets, the dataset size was not promising to achieve high accuracy in model training. In the beginning, the dataset size was 13100 in terms of the number of training images where each of the 10 classes consists of 1310 images. But face images carry plenty of information, to be specific, features that need to be extracted properly. Moreover, CNN is a data-driven approach and so model accuracy shares a positive correlation with the size of the dataset. Keras offers a function, data augmentation, allows researchers to extend their dataset. Data augmentation basically applies some basic operations like shear(10%), zoom(10%), rotate(30%), horizontal flip and generate additional image data from the existing dataset [26].

After applying data augmentation on the existing dataset, the dataset achieves a higher number of images, in other words, the dataset is augmented. The extended dataset includes a total of 39300 images and each class has 3930 images. That means dataset images are equally distributed among the classes to maintain equal or nearly equal recognition rates in all classes. Each input image has a dimension of 48×48. During the

TABLE II: Belief Rule Base for BRBES

Rule ID	Rule Weight	IF	THEN (Overall Mental State)		
		Person's Expression	Good	Normal	Bad
1	1	Angry	0.00	0.10	0.90
1	1	Disgust	0.00	0.15	0.85
1	1	Fear	0.00	0.20	0.80
1	1	Happy	1.00	0.00	0.00
1	1	Mockery	0.00	0.35	0.65
1	1	Neutral	0.50	0.50	0.00
1	1	Sad	0.00	0.25	0.75
1	1	Surprise	0.40	0.60	0.00
1	1	Think	0.30	0.70	0.00
1	1	Wink	0.00	0.55	0.45

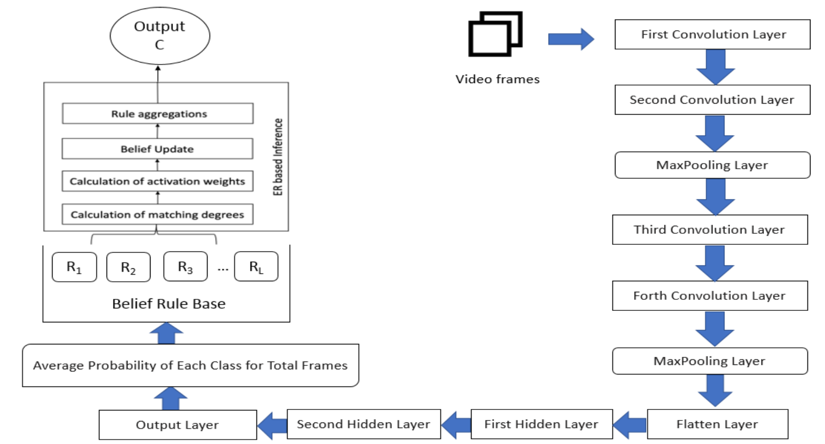

Fig. 2: Working Process of the Integration of CNN and BRBES

learning phase, 80% of the images have been selected for model training and the remaining 20% is for system validation. Apart from CNN, several machine learning models are taken into account for this facial expression classification task. Support Vector Machine and Random Forest model are applied to the dataset to observe the effectiveness of these models in facial expression detection. In this research, HOG features that represent the low-level features are extracted from the images of the dataset and fed into both SVM, random forest model, and track the performance. Moreover, more experiments have been conducted, providing our preprocessed data images as the learning input parameter to these machine learning models, and draw an overall performance comparison among all cases.

VIII. Result and Discussion

To illustrate the performance of the proposed model, both accuracy and loss graphs are shown in Fig. 3. Here x-axis refers to the number of model training epochs which refer to the number of training cycles through the full dataset and y-axis indicates the loss and accuracy respectively. If the accuracy graph is observed closely, in the beginning, validation accuracy is higher than the training accuracy for some epochs. Both the validation and training accuracy curves follow an upward trend as the number of epochs increases. In fact, dramatic growth takes place in the period between 1 to 5 epochs. After that growth rate diminishes firmly and stopped by the callback named EarlyStopping, offered by Keras, after 30 epochs. After the end of the training, the model achieves 98.28% training accuracy and 90.93% validation accuracy. The difference between these two accuracies is not significant. That expresses that the proposed model is not biased to training images rather it performs almost equally well to classify unseen images also. On the other hand, Fig. 4 shows the confusion matrix of the proposed model that describes how many images are correctly classified among the test images. Out of 10 classes, six classes cross the 90 (in percent) margin and the other four classes are around or cross the 80 (in percent) margin. Disgust class has the lowest recognition rate which is approximately 0.80. Actually, the trained model sometimes confuses it with angry expression because these two facial expressions share similar visual information. Fig. 5 depicts the real-time validation of the proposed integrated CNN-BRB model and it shows the overall mental state with 'Good', 'Normal' and 'Bad' referential value for respective facial expression video. The performance of the proposed model expresses the robustness while assessing the mental state of a person from facial expression. Table III depicts a overall performance comparison of the applied models. Validation accuracy of SVM is slightly improved from 0.67 to 0.69 when HOG features are passed as learning materials to the it rather than passing raw images. But, there is the opposite

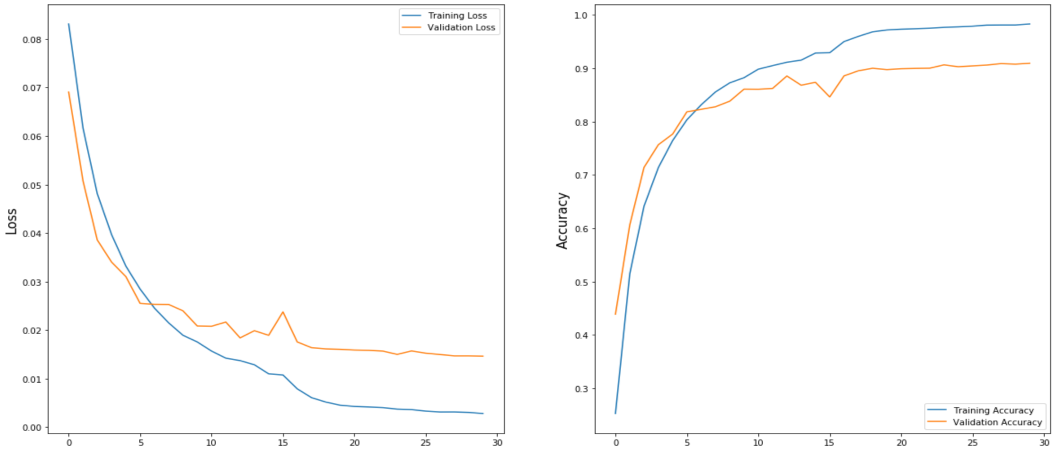

Fig. 3: Loss and accuracy graph of the proposed model

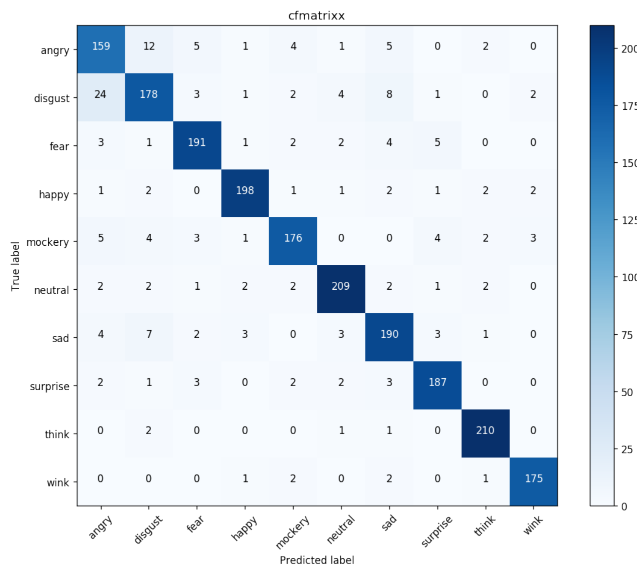

Fig. 4: Confusion matrix of the proposed model

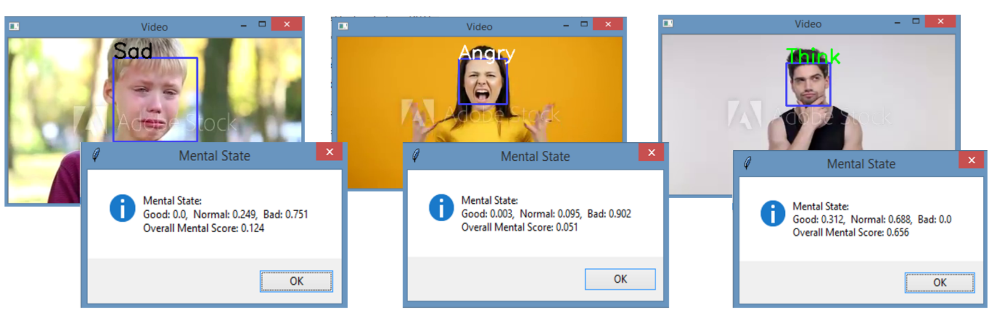

Fig. 5: Real time validation of the proposed integrated model

case for random forest model since it delivers higher prediction accuracy which is 0.73 by considering raw images rather than HOG features for training input. However, when it comes to CNN, it outperforms the other machine learning models with a proper margin in terms of all performance evaluation metrics: validation accuracy (0.90), recall (0.895), precision (0.915), f1 score (0.905). The formula for model accuracy are as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Here TP, FP, TN, FN symbolize True Positive, False Positive, True Negative and False Negative respectively.

TABLE III: Models' Performance Comparison Chart

Model Name	Validation Accuracy	Recall	Precision	F1 Score
SVM	0.67	0.674	0.673	0.672
HOG features + SVM	0.69	0.693	0.688	0.690
Random Forest	0.73	0.727	0.726	0.724
CNN	**0.90**	**0.895**	**0.915**	**0.905**

IX. Conclusion

The objective of this research is to assess the mental condition of an individual from ten facial expressions which are angry, disgust, fear, happy, mockery, neutral, sad, surprise, think and wink from video streaming data using both data-driven (CNN) and knowledge-driven (BRB) approach. The deep learning method is employed to generate class wise probabilities of a video-frame, and BRB is involved to use those class probabilities to trigger a particular rule for assessing the mental state. As this research deals with image data, CNN is taken into account as a deep learning method for its efficiency to handle image data. Although the recognition rate for each class of the CNN part of the proposed model is promising, efficiency can be improved in recognizing angry and disgust, which is under consideration as one of the future improvements. Pain assessment, patient behavior observation, autism identification, crime investigation, IoT based applications involving human behavior information are some of the cases where the proposed model can be beneficial.

References

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[2] T. U. Ahmed, S. Hossain, M. S. Hossain, R. ul Islam, and K. Andersson, "Facial expression recognition using convolutional neural network with data augmentation," in *2019 Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR)*. IEEE, 2019, pp. 336–341.

[3] T. U. Ahmed, M. S. Hossain, M. J. Alam, and K. Andersson, "An integrated cnn-rnn framework to assess road crack," in *2019 22nd International Conference on Computer and Information Technology (ICCIT)*. IEEE, 2019, pp. 1–6.

[4] M. S. Hossain, S. Rahaman, A.-L. Kor, K. Andersson, and C. Pattinson, "A belief rule based expert system for datacenter pue prediction under uncertainty," *IEEE Transactions on Sustainable Computing*, vol. 2, no. 2, pp. 140–153, 2017.

[5] N. Zeng, H. Zhang, B. Song, W. Liu, Y. Li, and A. M. Dobaie, "Facial expression recognition via learning deep sparse autoencoders," *Neurocomputing*, vol. 273, pp. 643–649, 2018.

[6] A. T. Lopes, E. de Aguiar, A. F. De Souza, and T. Oliveira-Santos, "Facial expression recognition with convolutional neural networks: coping with few data and the training sample order," *Pattern Recognition*, vol. 61, pp. 610–628, 2017.

[7] E. Owusu, Y. Zhan, and Q. R. Mao, "A neural-adaboost based facial expression recognition system," *Expert Systems with Applications*, vol. 41, no. 7, pp. 3383–3390, 2014.

[8] A. R. Rivera, J. R. Castillo, and O. O. Chae, "Local directional number pattern for face analysis: Face and expression recognition," *IEEE transactions on image processing*, vol. 22, no. 5, pp. 1740–1752, 2012.

[9] Y. LeCun, Y. Bengio *et al.*, "Convolutional networks for images, speech, and time series," *The handbook of brain theory and neural networks*, vol. 3361, no. 10, p. 1995, 1995.

[10] L. Luo, Y. Xiong, Y. Liu, and X. Sun, "Adaptive gradient methods with dynamic bound of learning rate," in *Proceedings of the 7th International Conference on Learning Representations*, New Orleans, Louisiana, May 2019.

[11] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. IEEE, 2010, pp. 94–101.

[12] I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee *et al.*, "Challenges in representation learning: A report on three machine learning contests," in *International Conference on Neural Information Processing*. Springer, 2013, pp. 117–124.

[13] N. Aifanti, C. Papachristou, and A. Delopoulos, "The mug facial expression database," in *11th International Workshop on Image Analysis for Multimedia Interactive Services WIAMIS 10*. IEEE, 2010, pp. 1–4.

[14] O. Langner, R. Dotsch, G. Bijlstra, D. H. Wigboldus, S. T. Hawk, and A. Van Knippenberg, "Presentation and validation of the radboud faces database," *Cognition and emotion*, vol. 24, no. 8, pp. 1377–1388, 2010.

[15] P. I. Wilson and J. Fernandez, "Facial feature detection using haar classifiers," *Journal of Computing Sciences in Colleges*, vol. 21, no. 4, pp. 127–133, 2006.

[16] M. R. Sabbir Hridoy, M. S. Hossain, R. U. Islam, and K. Andersson, "A web based belief rule based expert system for assessing flood risk," in *Proceedings of the 19th International Conference on Information Integration and Web-based Applications & Services*. ACM, 2017, pp. 434–440.

[17] R. Karim, K. Andersson, M. S. Hossain, M. J. Uddin, and M. P. Meah, "A belief rule based expert system to assess clinical bronchopneumonia suspicion," in *2016 Future Technologies Conference (FTC)*. IEEE, 2016, pp. 655–660.

[18] M. N. Jamil, M. S. Hossain, R. Ul Islam, and K. Andersson, "A belief rule based expert system for evaluating technological innovation capability of high-tech firms under uncertainty," in *Joint 2019 8th International Conference on Informatics, Electronics & Vision (ICIEV)*. IEEE, 2019.

[19] R. Ul Islam, K. Andersson, and M. S. Hossain, "A web based belief rule based expert system to predict flood," in *Proceedings of the 17th International conference on information integration and web-based applications & services*, 2015, pp. 1–8.

[20] M. S. Hossain, I. B. Habib, and K. Andersson, "A belief rule based expert system to diagnose dengue fever under uncertainty," in *2017 Computing Conference*. IEEE, 2017, pp. 179–186.

[21] M. S. Hossain, S. Rahaman, R. Mustafa, and K. Andersson, "A belief rule-based expert system to assess suspicion of acute coronary syndrome (ACS) under uncertainty," *Soft Computing*, vol. 22, no. 22, pp. 7571–7586, 2018.

[22] T. Mahmud and M. S. Hossain, "An evidential reasoning-based decision support system to support house hunting," *International Journal of Computer Applications*, vol. 57, no. 21, pp. 51–58, 2012.

[23] M. S. Hossain, M. S. Khalid, S. Akter, and S. Dey, "A belief rule-based expert system to diagnose influenza," in *2014 9th International Forum on Strategic Technology (IFOST)*. IEEE, 2014, pp. 113–116.

[24] M. S. Hossain, P.-O. Zander, M. S. Kamal, and L. Chowdhury, "Belief-rule-based expert systems for evaluation of e-government: a case study," *Expert Systems*, vol. 32, no. 5, pp. 563–577, 2015.

[25] M. S. Hossain, F. Ahmed, F. Johora, and K. Andersson, "A belief rule based expert system to assess tuberculosis under uncertainty," *Journal of Medical Systems*, vol. 41, no. 43, pp. 1–11, 2017.

[26] L. Perez and J. Wang, "The effectiveness of data augmentation in image classification using deep learning," *arXiv preprint arXiv:1712.04621*, 2017.