Joint 2019 8th International Conference on Informatics, Electronics & Vision (ICIEV) &
3rd International Conference on Imaging, Vision & Pattern Recognition (IVPR)

Facial Expression Recognition using Convolutional Neural Network with Data Augmentation

Tawsin Uddin Ahmed*, Sazzad Hossain†, Mohammad Shahadat Hossain‡, Raihan Ul Islam§ and Karl Andersson¶
*Department of Computer Science and Engineering, University of Chittagong, Chittagong, Bangladesh
Email: tawsin.uddin@gmail.com
†Department of Computer Science and Engineering, University of Liberal Arts Bangladesh, Dhaka, Bangladesh
Email: sazzad.hossain@ulab.edu.bd
‡Department of Computer Science and Engineering, University of Chittagong, Chittagong, Bangladesh
Email: hossain_ms@cu.ac.bd
§Department of Computer Science, Electrical and Space Engineering, Luleå University of Technology, Skellefteå, Sweden
Email: raihan.ul.islam@ltu.se
¶Department of Computer Science, Electrical and Space Engineering, Luleå University of Technology, Skellefteå, Sweden
Email: karl.andersson@ltu.se

***Abstract*—Detecting emotion from facial expression has become an urgent need because of its immense applications in artificial intelligence such as human-computer collaboration, data-driven animation, human-robot communication etc. Since it is a demanding and interesting problem in computer vision, several works had been conducted regarding this topic. The objective of this research is to develop a facial expression recognition system based on convolutional neural network with data augmentation. This approach enables to classify seven basic emotions consist of angry, disgust, fear, happy, neutral, sad and surprise from image data. Convolutional neural network with data augmentation leads to higher validation accuracy than the other existing models (which is 96.24%) as well as helps to overcome their limitations.**

***Keywords*—Convolutional neural network, data augmentation, validation accuracy, emotion detection.**

I. Introduction

Facial expression refers to the movement of facial muscles that carry emotions expressed by a person. It provides information about the mental status of that person. Emotion is a mental condition that the person goes through. It is what he feels inside as the reaction of the events that take place around him. Information about a person's mental condition lies in his facial condition in many cases. Analysis of facial expression has many applications such as in lie detectors, robotics and in art [1]. Improvement in the skill of facial expression recognition is required for an intelligent agent to communicate with human as a part of machine-human collaboration as well as with robots as a part of robot-robot interaction [2].

As research on facial expression recognition has been conducting for years, research progress on this topic is commendable [3] [4]. Fluctuation in recognition rate among the classes is one of the issues for most of the research as they have lower recognition rate to detect emotions like disgust and fear [5] [6].

The purpose of this research is to develop a facial expression recognition system which can classify an image into seven different classes of emotion. In addition, the improvement of the validation accuracy compared to the other existing systems as well as to maintain commendable and equal or nearly equal recognition rate for each class will also be addressed.

Convolutional neural network in facial expression recognition has been applied in a few research but inconsistency in recognition rate among classes is one of the issues for most of the research as they have lower recognition rate in disgust and fear [5] [6] [7]. We come up with the idea of CNN with data augmentation and combined dataset collected from several datasets which leads this research to higher validation accuracy as well as higher and nearly equal recognition rates compared to the existing models.

The remaining sections of this article consist of: related work on facial expression recognition, an overview of the methodology of this research, data collection and preprocessing, experiment with data augmentation, how the proposed system has been implemented, result and discussion, conclusion and future work.

II. Related Work

Research on facial expression has been conducting for years. But there was always a room for improvement for every research. That is why there are many opportunities regarding this topic.
In [5] the main goal of their research is to improve accuracy of a particular dataset FER2013. They applied convolutional neural network as the methodology of their proposed model to classify seven basic emotions. This research demonstrated the success of convolutional neural network to improve the accuracy of biometric applications. However, there exists fluctuation in recognition rate of each class as they could not maintain the equal or nearly equal recognition rate for each

978-1-7281-0788-2/19/1.00

class. Although overall accuracy has been achieved at 91.12%, recognition rate in classifying disgust and fear only stands at 45% and 41% respectively.

In [6] earlier before this, researchers had developed facial expression recognition system based on posed images in static environment. However, [6] introduced a facial expression dataset named RAF-DB that consists of images of different ages and poses in dynamic environment. They applied deep locality preserving CNN method to classify 7 basic emotions. Their proposed model was trained based on RAF-DB and CK+ datasets. Although 95.78% of accuracy had been achieved, recognition rate in disgust and fear only stands at 62.16% and 51.25% respectively.

In [7] they have mentioned this as the extension of their previous work. To classify the six basic emotions they have applied deep convolutional neural network which is a combination of convolutional neural network coupled with the deep residual blocks. They have trained their model on two datasets named Extended Cohn Kanade (CK+) and Japanese Female Facial Expression (JAFFE). Better performance than state-of-the-art approach and higher accuracy than other models have been considered as their research success. Accuracy has been achieved at 95.24%. As their system is based on two datasets, it is biased to those datasets. In addition, they could not classify emotions from the image that carry geometrically displaced faces.

In [8] they have compared two types of facial feature extraction methods. One is geometric positions of fiducial points and the other is Gabor-wavelet coefficient fetch method. As a result of this comparison, they have shown that the gabor-wavelet coefficient fetch method performs better than the other one. They have succeeded to find out the number of hidden layers required which is five to seven in order to achieve higher recognition rate. Accuracy has been achieved at 90.1%. Although they have not shown individual class recognition rate, admitted having less recognition rate in fear class.

III. Methodology

Convolutional Neural Network is considered as the methodology that is used with data augmentation in this research. Dataset that is used in this research has variation as data was collected from different datasets. As a result, the proposed model is not biased to any particular dataset. The event flow chart of this system is illustrated in fig. 1.

In figure 1, at first, the model takes an image from the dataset

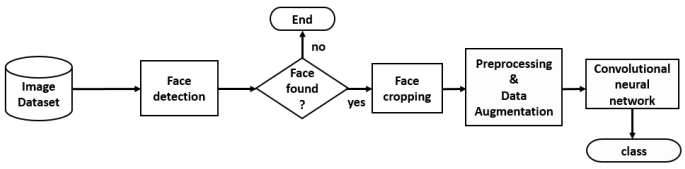

Fig. 1. System Flow Chart

and detects face from the image by Cascade Classifier. If face is found, then it is sent for preprocessing. Data have been augmented by ImageDataGenerator function offered by the Keras API. At last, the augmented dataset is fed into CNN in order to predict the class.

The model that is used to classify the facial expression contains 3 convolution layers with 32, 64 and 128 filters respectively and the kernel size is 3x3.

Convolution over an image f(x, y) using a filter w(x, y) is defined in equation (1):

$$w(x,y) * f(x,y) = \sum_{s=-a}^{a} \sum_{t=-b}^{b} w(s,t) f(x-s, y-t) \quad (1)$$

The activation function that has been used in convolution layer is Relu activation function. Relu is applied to introduce the non-linearity of a model [9] and it is shown in equation (2):

$$f(x) = max(0, x) \quad (2)$$

Model has been provided with 48X48 sized images as model input. The input shape of the model is (48,48,1), where 1 refers to the number of channels exists in input images. Images have been converted into grayscale that is why the number of channels is 1. After convolution layer, the model has 2*2 pool size pooling layer and max pooling has been chosen. Next, there are four fully connected layers which consist of 750, 850, 850 and 750 nodes respectively. Like convolution layer,

TABLE I
System Architecture

Model Content	Details
First Convolution Layer	32 filters of size 3x3,ReLU, input size 48x48
First Max Pooling Layer	Pooling Size 2x2
Second Convolution Layer	64 filters of size 3x3, ReLU
Second Max Pooling Layer	Pooling size 2x2
Third Convolution Layer	128 filters of size 3x3, ReLU
Third Max Pooling Layer	Pooling size 2x2
First Fully Connected Layer	750 nodes, ReLU
Dropout Layer	Excludes 50% neurons randomly
Second Fully Connected Layer	850 nodes, ReLU
Dropout Layer	Excludes 50% neurons randomly
Third Fully Connected Layer	850 nodes, ReLU
Dropout Layer	Excludes 50% neurons randomly
Forth Fully Connected Layer	750 nodes, ReLU
Dropout Layer	Excludes 50% neurons randomly
Output Layer	7 nodes for 7 classes, SoftMax
Optimization Function	Stochastic Gradient Descent (SGD)
Learning Rate	0.01
Callback	EarlyStopping,ReduceLROnPlateau, ModelCheckpoint,TensorBoard

Relu activation function has been applied in hidden layers. Right after each hidden layer, a dropout layer has been inserted and the value of dropout has been set to 0.5. It randomly deactivates 50% nodes from the hidden layer to avoid over-fitting [10]. At last, the output layer of the model consists of 7 nodes as it has 7 classes. Softmax has been used as activation function in the output layer.

$$Softmax(x) = \frac{e^j}{\sum_i e^i} \quad (3)$$

As model optimizer Stochastic Gradient Descent (SGD) [11] has been used with learning rate 0.01. As loss function Categorical Crossentropy has been used. Callbacks which have been included in the model are EarlyStopping, ReduceLROnPlateau, ModelCheckPoint and TensorBoard.
The overview of the convolutional neural network architecture that has been designed for this research is given in Table I.

IV. DATA COLLECTION AND PREPROCESSING

Datasets have been collected from different sources so that the output is not biased towards a particular dataset. Various standard facial datasets are available online:

- CK and CK+ [12]
- FER2013 [13]
- The MUG Facial Expression Database [14]
- KDEF & AKDEF [15]
- KinFaceW-I and II [16]

It is worth mentioning that FER2013 dataset has been modified as it contains many wrongly classified images which causes lower accuracy gain by the previous research that is based on this dataset [5]. Dataset samples of the model are shown in fig 2.
For data preprocessing following steps are considered:

Fig. 2. Dataset Samples

a. Face detection and crop
b. Grayscale conversion
c. Image normalization
d. Image augmentation

A. Face detection and crop

The process of face detection which is called Face Registration is a process of detecting face location from an image. OpenCV Cascade classifier [17] has been used to detect face from the images. After detecting the face, the face portion has been cropped out to avoid background complexity so that the model training becomes more efficient.

B. Grayscale conversion

Images have been resized into 48*48 pixels having 3 channels red, green and blue. To reduce the complexity in pixel values, dataset images have been converted into grayscale having only one channel [18]. So it has become pretty much easy for the model to learn.

C. Image normalization

Normalization has been applied to model dataset which is a process that modifies the range of pixel intensity values to a certain limit. It is a process by which contrast or histogram of the images can be stretched so that it enables deep network to analyze the images in a better way [19].

Fig. 3. Data Preprocessing

D. Image augmentation

Convolutional neural network requires a large number of datasets in order to get better performance. If the dataset is large, it can extract more features from them and match with the unlabeled data. If it is not possible to collect enough data, data augmentation could be an option to improve the performance of the model. Image augmentation generates additional images by applying some operations on existing image dataset, such as random rotation, shifts, shear, flips etc.

V. EXPERIMENT WITH DATA AUGMENTATION

Keras API facilitates data augmentation process by introducing ImageDataGenerator function which through several operations can be applied on the existing dataset to generate more new data. As the parameters for ImageDataGenerator function, five operations have been included which are rotation at a certain angle, shearing, zooming, horizontal flip, rescale. The parameters with respective values are shown in Table II.

TABLE II
DATA AUGMENTATION PARAMETERS

Operation Type	Value
Horizontal Flip	True
Rotation	0.30
Rescale	1./255
Shear	0.20
Zoom	0.20

Before data augmentation, the dataset had a total of 12,040 images. Each class contains around 1720 images. As CNN is a data-driven approach, in order to achieve more improved model performance it had been decided to enrich the existing dataset with more images. So, some operations like zoom,

rotation, shear, flip and position shift in certain position have been applied to the existing dataset so that more new data can be generated. After applying data augmentation to the dataset,

Fig. 4. Data Augmentation

it has got a total of 36,120 images having around 5160 images per class. Image or data augmentation is being used to improve deep learning in image classification problem [20]. Therefore, including data augmentation techniques in facial expression recognition has been chosen for this research. During learning process, 80% of the images have been selected for model training and the remaining 20% is for system validation.
For more experiment, splitting ratio was changed in a more challenging way to test the performance of the proposed model. 65% of the dataset was selected for training purpose and the remaining 35% was chosen for testing so that it could be verified if the model performs well with larger dataset.

VI. SYSTEM IMPLEMENTATION

The program has been written in python programming language, using the Spyder IDE. The libraries required in this experiment are Keras, Tensorflow [21], numpy, PIL, OpenCV and matplotlib. Tensorflow was used as system backend whereas keras helped the system by providing built-in functions like activation functions, optimizers, layers etc. OpenCV was mainly used for image preprocessing such as face detection (Cascade Classifier), grayscale conversion, image normalization. Data augmentation was performed by keras API. Matplotlib has been used to generate confusion matrix. A Graphical User Interface (GUI) to display outputs of the proposed model has been built using HTML, CSS, JavaScript. The saved Neural Network model has been hosted on a Python Flask server and it allows a user to choose an image from the local device as input. When the user clicks on the "Predict" button the class of the image is shown after execution. Some sample screenshots of the graphical user interface are given in fig 5: Real-time images have been provided as system input so that it can be shown that the model has the ability to predict unseen images accurately. It should be mentioned that whenever a user provides an image as input in the system, it preprocesses the image in the same way when the model has been trained. That means at the beginning whenever an image of an arbitrary size is given by the user, the system converts it to 48*48 sized image. Then with the help of Cascade Classifier [17], the model detects the face from the image. It is mainly

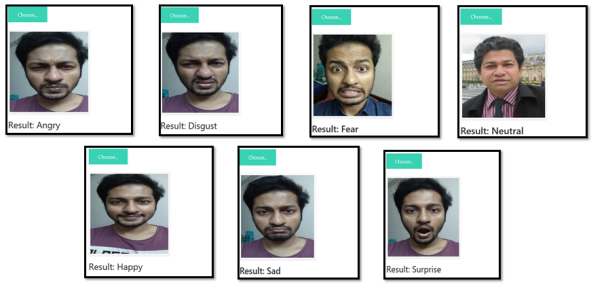

Fig. 5. Real Time Validation

the region of interest which is been cropped afterward. As the model has been trained on grayscale images, the system converts the rgb image that contains 3 channels red, green and blue to gray image which consists of only 1 channel. Then to ease the classification task the system has applied image normalization on the image. Then it is sent to the customized Convolutional Neural Network for classification.

VII. RESULT AND DISCUSSION

Even though the proposed model has been trained on combined dataset, it has been successful to achieve validation accuracy of 96.24%. This model has succeeded to maintain higher and nearly equal recognition rate for each class as well as it can classify geometrically displaced face images. The

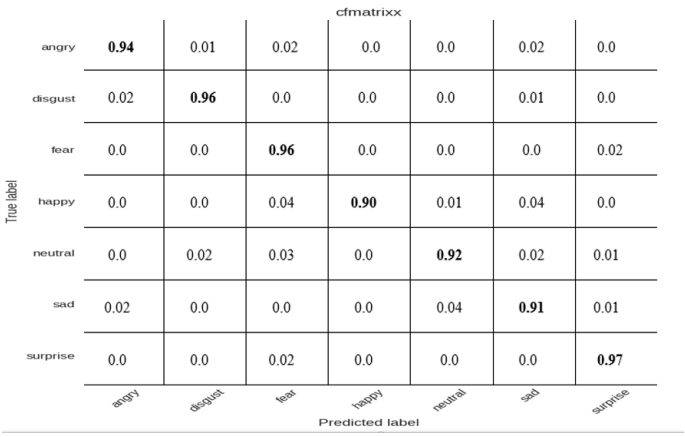

Fig. 6. Confusion Matrix

proposed system delivers a firm classification output. Though there exists a little fluctuation in recognition rate among the seven classes, it is still better compared to the other existing models. [7] could not predict geometrically displaced face images and [8] had low recognition rate in fear class. This research has overcome these limitations. The proposed model, convolutional neural network with data augmentation, has been successful to achieve validation accuracy of 96.24% which is the highest accuracy so far in facial expression recognition. This model has succeeded to maintain higher and nearly equal recognition rate for each class as well as it can classify geometrically displaced face images. Tensorboard has been used to visualize model validation accuracy progress over time.

In fig 7, x-axis refers to the number of epochs and y-axis

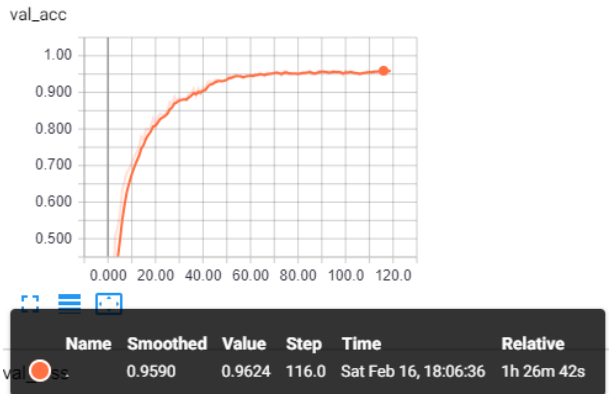

Fig. 7. Real Time Evaluation of our Model Training

refers to the recognition rate. It can be observed that our desired accuracy is achieved after only 120 epochs. Data or image augmentation boosted the model in terms of accuracy. It is proved that data augmentation performs a major role in efficient model development if the dataset is not very large.

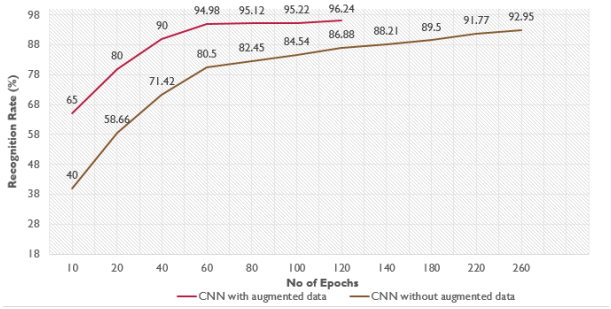

Fig. 8. Training curve before and after data augmentation

Fig 8 shows the situation before and after data augmentation. It can be noticed that CNN with data augmentation has got accuracy of 96.24% after only 120 epochs whereas CNN without data augmentation has required 260 epoch to get 92.95% validation accuracy. It should be mentioned that the new images generated from the existing image dataset by data augmentation are uniquely identified which means image variation exists in the dataset. Along with this, dropout and EarlyStopping callbacks have been used so that model overfitting can be avoided. Moreover, model performance has been evaluated with 65% and 35% splitting ratio for training and testing. It has performed well even with larger test data as the accuracy has been achieved at 95.87%. Though accuracy is slightly lower then the previous accuracy (with 80:20 splitting ratio), it is fair because the model was being tested with larger test data.

VIII. CONCLUSION AND FUTURE WORK

In this research work, the main agenda was to find out the improvement opportunities for the existing facial expression recognition system. Finding out their limitations and applying probable solutions to overcome the limitations were the objectives of this research. The Convolutional Neural Network method with data augmentation has been proved to be more efficient compared to other machine learning approaches in case of image processing [20]. The proposed model has achieved higher validation accuracy than any other existing model. The Graphical User Interface allows users to do real-time validation of the system. We have considered seven discrete and unique emotion classes (angry, disgust, fear, happy, neutral, sad and surprise) for emotion classification. So, there is no overlapping among classes. However, we are planning to work with compound emotion classes such as surprised with happiness, surprised with anger, sadness with anger, surprised with sadness and so on. In addition, an aggregated view of the facial expression by combining different emotions as well as compound emotions will be determined under uncertainty by using sophisticated methodology like Belief Rule Based Expert Systems (BRBES) in an integrated framework [22] [23] [24] [25] [26]. As different problems would require different network architectures it is required to figure out which architecture is the best for a particular problem. Though the proposed model has achieved a commendable result, it needs some improvements in some areas like:

- Adding more data in each class in order to get more accurate result as it is known that deep learning is a data-driven approach.
- Getting higher recognition rate in happy.

In the future, researchers can try to develop the model more efficiently so that a more standard facial expression recognition system can be delivered.

IX. ACKNOWLEDGMENT

This study was funded by the Swedish Research Council under grant 2014-4251.

REFERENCES

[1] M. Shidujaman, S. Zhang, R. Elder, and H. Mi, ""roboquin": A mannequin robot with natural humanoid movements," in *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2018, pp. 1051–1056.

[2] M. Shidujaman and H. Mi, ""which country are you from?" a cross-cultural study on greeting interaction design for social robots," in *International Conference on Cross-Cultural Design*. Springer, 2018, pp. 362–374.

[3] M. Pantic and L. J. Rothkrantz, "Automatic analysis of facial expressions: The state of the art," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 12, pp. 1424–1445, 2000.

[4] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," *IEEE transactions on pattern analysis and machine intelligence*, vol. 31, no. 1, pp. 39–58, 2009.

[5] N. Christou and N. Kanojiya, "Human facial expression recognition with convolution neural networks," in *Third International Congress on Information and Communication Technology*. Springer, 2019, pp. 539–545.

[6] S. Li and W. Deng, "Reliable crowdsourcing and deep locality-preserving learning for unconstrained facial expression recognition," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 356–370, 2019.

[7] D. K. Jain, P. Shamsolmoali, and P. Sehdev, "Extended deep neural network for facial emotion recognition," *Pattern Recognition Letters*, 2019.

[8] Z. Zhang, M. Lyons, M. Schuster, and S. Akamatsu, "Comparison between geometry-based and gabor-wavelets-based facial expression recognition using multi-layer perceptron," in *Proceedings Third IEEE International Conference on Automatic face and gesture recognition*. IEEE, 1998, pp. 454–459.

[9] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, 2011, pp. 315–323.

[10] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[11] L. Bottou, "Large-scale machine learning with stochastic gradient descent," in *Proceedings of COMPSTAT'2010*. Springer, 2010, pp. 177–186.

[12] 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.

[13] P. Giannopoulos, I. Perikos, and I. Hatzilygeroudis, "Deep learning approaches for facial emotion recognition: A case study on fer-2013," in *Advances in Hybridization of Intelligent Methods*. Springer, 2018, pp. 1–16.

[14] 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.

[15] M. G. Calvo and D. Lundqvist, "Facial expressions of emotion (kdef): Identification under different display-duration conditions," *Behavior research methods*, vol. 40, no. 1, pp. 109–115, 2008.

[16] M. Shao, S. Xia, and Y. Fu, "Genealogical face recognition based on ub kinface database," in *CVPR 2011 WORKSHOPS*. IEEE, 2011, pp. 60–65.

[17] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *null*. IEEE, 2001, p. 511.

[18] M. Grundland and N. A. Dodgson, "Decolorize: Fast, contrast enhancing, color to grayscale conversion," *Pattern Recognition*, vol. 40, no. 11, pp. 2891–2896, 2007.

[19] R. C. Gonzalez, "Digital image processing/richard e," *Woods. Interscience, NY*, 2001.

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

[21] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard *et al.*, "Tensorflow: A system for large-scale machine learning," in *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*, 2016, pp. 265–283.

[22] 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.

[23] 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*. ACM, 2015, p. 3.

[24] 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.

[25] 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.

[26] M. S. Hossain, K. Andersson, and S. Naznin, "A belief rule based expert system to diagnose measles under uncertainty," in *World Congress in Computer Science, Computer Engineering, and Applied Computing (WORLDCOMP'15): The 2015 International Conference on Health Informatics and Medical Systems 27/07/2015-30/07/2015*. CSREA Press, 2015, pp. 17–23.