ISSN 2278-3091
Volume 9 No.2, March -April 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse65922020.pdf
https://doi.org/10.30534/ijatcse/2020/65922020

Age and Gender prediction using Convolution, ResNet50 and Inception ResNetV2

Avuthu Sai Meghana[1], Sudhakar S[2], Arumugam G[3], Srinivasan P[4], Kolla Bhanu Prakash[5]

[1]Dept. of CSE, Koneru Lakshmaiah Education Foundation, India,
asmeghana97@gmail.com
[2]Dept. of CSE, Sree Sakthi Engineering College, Coimbatore, Tamil Nadu India,
sudhasengan@gmail.com
[3]Faculty of Electrical and Electronics Engineering, Al Shabaka Technical Institute, Dubai,
arumugam@astidubai.ac.ae
[4]Dept. of ECE, Sona College of Technology, Salem, Tamil Nadu, India,
srinisamy2004@gmail.com
[5]Dept. of CSE, Koneru Lakshmaiah Education Foundation, India,
drkbp@kluniversity.in

ABSTRACT

Age and gender prediction are used extensively in the field of computer vision for surveillance. Advancement in computer vision makes this prediction even more practical and open to all, thus enables the world to come up with datasets, one of which, used in this paper, is UTKFace that has 1000 pictures of male and female actors ageing from 0 to 100. In this paper, we propose a Convolution Neural Network (CNN) with ResNet50 architecture to predict age and gender. CNN is a Neural Network (NN) algorithm that extracts the deep features from the image and specifies the desired output at the final layers. Age prediction is approximately near to the real values with a five difference in both ways. Gender prediction is accurate in all the test data presented to the model. Validating with arguments shows no change in training and validation. Our model successfully executed with approximately 80% in gender prediction and 60% in age prediction that can be furtherly advanced with pipelining with other classification models and much larger real-world datasets.

Key words: CNN, ResNet50, NN, UTKFace

1. INTRODUCTION

As technology increases, the applications that combine the advanced fields of pattern recognition and image processing are used to find age and gender. In today's world, age plays a prominent role, when you appear for an interview, health check-ups. The information of age is used in many government, private and advertising sector organization to find the culprits, employee eligible for the job, audience to be targeted for their publicity of product respectively. However, it's not that easy to find the age of a person, and there are constraints that restrict us from seeing the correct age among the set of images. Human face contains lot of information through their expressions. These expressions confuse us while finding the age, and as their expression change, the facial feature differs, resulting in either higher or low than the people's ideal age.

The age estimation plays a prominent role in the applications like biometric evaluation, 3d face marking, virtual makeup, and virtual try-on applications for jewelry and eye-ware by mapping the face according to the age found. Lens kart is such an application that gives the try-on option for their customers. Age estimation is a subfield of face recognition and face tracking which in combination can predict the health of the individual. Many health care applications use this mechanism to keep track of health by monitoring their daily activities. China uses this face detection technique in service driver identification and jaywalker identification. Some other countries use it for worshipper identification and advertising etc. [1]. Finding the correct dataset for training the model is a crucial task. Since the real-time data is massive, the computation and the time to prepare the model are high. It's been a tough task after implementing several methods from machine learning, and the accuracy increases drastically [1]. By eliminating the barrier of expressions, we have the possibility to find the best features leading to an accurate measure of gender and age.

To predict the age and gender, we use a vital range of machine learning and deep learning algorithms. CNN (convolution neural network) is one of the most used techniques for age and gender detection. In this paper, we use open cv and CNN to predict the age and gender of any given person's image.

2. RELATED WORK

Felix Anda, David Lillis, and others who were part of the forensic and security research group have led their research towards the evaluation of the state of art behavior of different cloud-based biometric services (CBBS). The Model-View-Controller (MVC) is used to collect the data according to the criteria provided. The unnecessary and wrong images are removed at the manual filtering step. The mechanism used for evaluation to find the least mean square error (MAE) among different CBBS and pre-trained models is an empirical evaluation based on observation. Here the Caffe model is used in sensing the age and gender [1].

Md. Hafizur Rahman, Md. Abul Bashar has identified the age and gender by converting the RBG (red, blue, green) image to YCbCr (luminance, chrominance blue, and red) format as it is used in video compression. The primary usage of this format is to find the skin-tone. The lighting compensation (LC) algorithm is used in the pre-processing phase to enhance the image and restore natural colours. Gabor filter is used for feature extraction from the pre-processed image and classified using the logistic regression [2].

In this paper, curvature changes in non-uniform Rational B-Splines (NURBS) deformations are used to detect the variation in age and gender-related surface patterns of MR images. By this, the changes in brain development corresponding to the age and gender are found [3].

A wide range of algorithms is used and compared at different stages in distinguishing age and gender. As a pre-processing technique, population pyramids scaling is used, which is then classified for gender detection according to the accuracy score between linear SVM and logistic regression. The age prediction is made in 2 methods, either using node attributes with the multinomial logistic process (MN Logistic) or by using network topology with a communication network structure and reaction-diffusion algorithm [4].

The spotting of age and gender is done by using a deep residual learning network with connections that contains a gender estimation network and a gender-specific age network where the output from gender network are used as weights for estimating the outputs of the two age networks. The specified model used regression for estimation of age [5].

The database FG-net has the predefined columns that specify the correct location of eyes, nose, and lips, etc. such that they can be identified from the set of images easily. This paper includes the usage of SVM for the classification of gender and age [6].

In order to the categorization of age and gender from the dataset, predefined information regarding the gender and age levels between 17 to 80. The three different feature vectors are considered, namely duration of fixation, density, arcs, which are considered while implementing the protocols to categorize the male and female from the given dataset and then an ADA-boost protocol in the pre-processing. The final methodology used is the gaze analysis technique for human identification (GANT) for prediction [7].

[8] Considers a dataset with different gender and age groups, where the personality, age, and gender labels are considered as a corpora. The open SMILE is used to extract features from an emotional challenge competition in 2009. These features are them split, and shared representation between the corpora is found by the shared-hidden-layer auto-encoder (SHLA). A support vector machine with a linear kernel is being used for classification, and later the cross-validation is done.

[9] uses the WC-CNN (integral component CNN), which is characterized into two types of neural networks, the whole face network, and the facial component network. The entire face network is used as the primary classifier, and the resulting classification is evaluated by the confidence analysis. Finally, the global and local network's outputs are joined to achieve the desired partitioning of age and gender.

Usage of deep networks VGG-16 architecture for feature extraction, and the age estimation is done using the fully connected layers (fc6). To train the dataset, structured output SVM (SO-SVM) is used. The RBF kernel is the key to gender and smile classification [10].

3. METHODOLOGY

This paper comprises of a CNN model that distinguishes age and gender from the UTK Face dataset. The dataset contains a considerable number of images ranging from newborn babies to 100 years person. The data is trained, and the testing is done on real-time images. The resNET50, which is a model of CNN, is used to prepare the images. Since the data is broad and computation and prediction with this kind of data will be hard and time-consuming. So, there is a requirement of the optimizer that is Adam optimizer. All the models use the TensorFlow platform in the background to generate the age and gender of the particular image.

3.1 Convolution Neural Network

The Neural network (NN) is a popular model in machine learning. It is developed based on the human brain's architecture with neurons. The neuron in the neural network has the functionality of taking the input and generates an output by applying the input on a function. These functions are entitled as activation functions. The neural network carries 3 layers namely input, hidden and output layers. Each layer in the neural network holds a bunch of neurons, the neurons in the input layer represent the number of features chosen. The aggregate amount of hidden layers depends upon the model, data size, and the complication in the problem. The neuron number in the hidden layer doesn't rely on the number of features, and the output of this is given to the logistic functions to observe the probability score of each class at the output layer. By the application of biases or weights in the hidden layer and activation, the feature gives a nonlinear network. The accurate prediction of weights to be applied at the hidden layer can be predicted by the backpropagation algorithm.

CNN is the most popular deep learning algorithm, used tremendously in computer vision. It is computationally efficient and instinctively spots prominent features without any supervision. CNN has the same layers compared with the traditional neural network, but the hidden layer consists of different kinds of internal layers, specifically the convolutional layer, pooling layer, fully connected layer, and normalization layer. CNN has a prominent role in computer vision and image processing, to process an image which is an array of pixels is taken in the form of a matrix. An image can be in color or black and white form that is RBG, grayscale, respectively. The image matrix that enters the convolution layer merges with the kernel or convolution filter that leads away for the feature map.

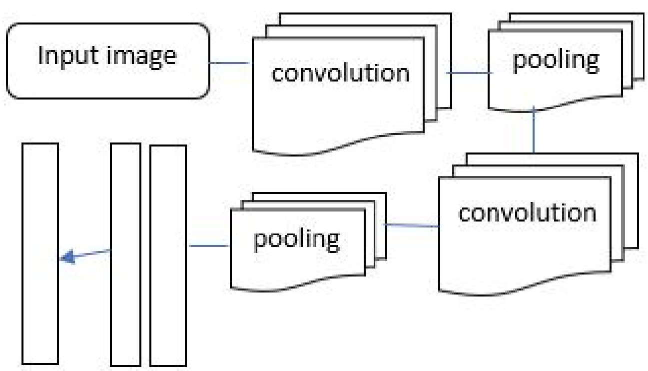

Figure 1: Architecture of 2 layered CNN

The convolution process includes the sliding of the filter over the input. According to the size of the filter, the input that coincides with the screen is defined as the receptive field [11]. The summation of the element-wise multiplication of the pixel values between input and the filter is added to as the first-pixel value in the feature map matrix of the size equal to the kernel. In the same way, pixel values in the feature map are filled by sliding the core over each pixel in the input image matrix. By the change in the filter, edge detection, sharpening, identity, and box blur operations can be performed on the input. The sliding of the filter in this layer is determined by the stride, and the future map is derived correctly by adding the zeros to the input image at edges to match the kernel, which is called padding.

After convolution, the at pooling is the concept where the dimensionality of data is reduced. The pooling layer reduces each feature map by three different types of pooling methods that are max pooling, average pooling, some pooling. In majority cases, max pooling is used in this layer. According to the pooling window size, the stride is divided. One of the methods mentioned above is applied, and the pixel of the window is filled. The pooling window size is 2x2 in many large numbers of cases.

The output of the pooling layer is flattered to generate a vector as a fully connected layer takes vectors as input. The process of flattering includes the conversion of a 3-dimensional matrix to a one-dimensional vector.

3.2 ResNet-50

Residual Network (ResNet-50) or residual neural network is a deep learning model which is a part of CNN used to classify images. It is mainly suitable for large datasets with more than 1000 images. It's a 50 layered deep network model which has about 23 million training parameters. Some of the similar training models are AlexNet, GoogleNet, and VGG19. The feature of ResNet-50 is that it has identical connections.

This model contains five stages with residual blocks containing three layers, each with the convolutions. As the traditional NN, this network with remaining blocks provides the output of one layer to the next layer and also provides by hoping 2-3 layers away. This process is named as identity connections.

3.3 Adam Optimizer

This optimization is used in place of stochastic gradient descent. The key features of the algorithm are computationally useful, less memory required, suitable for extensive data, easy to implement, used for noisy data.

3.4 Experimentation

We have used the UTKFace dataset for the identification of age and gender. On this dataset, we do pre-process, train the data, train the network, modeling, and evaluation [12].

3.4.1 Dataset

The UTKFace dataset has different images with a life span from 0 to 100 and around 10000 images and more [13]. Each image has close and clear facial features along with expressions, different poses, and the images are at various resolutions. This dataset provides the cropped version of the original images for better feature detection. Each image is labeled by its age, gender, along with its landmarks. Each image in the dataset is labeled with age, followed by gender, race, and date and time the image when it collected into the UTKFace dataset [14].

3.4.2 Pre-Processing

At this stage, all the images in the dataset are inspected to find out if there is an image file without the label in the format of its representation. If any such image file is found, then it is either removed/any estimated values are assigned. Since we have the cropped images along with the data set else, this process of cropping can be categorized under the pre-processing phase.

3.4.3 Training Data

In the UTKFace dataset, the cropped images don't have any borders or margins around the face, which is called as tightly crop. There is a need to be tightly crop [15] the images in the main images folder, which is accessed in the python program. To align and crop, we arrange the margin argument to 0 when the dataset is trained by taking the weights into consideration.

3.4.4 Modeling

For the sensing of age and gender through images, we used a neural network model called the ResNet50, which is a CNN. A single CNN [16] is used to find the age and gender. In this only CNN, both the assumptions are simultaneously predicted, such that there is no need to run two different models for different attributes. By the various stages of CNN [21] [22], the input images are first convoluted according to the filter, and later the required features are pooled out, and finally, by fully connected layer, the features and weights are mapped such that we could estimate the age and gender of the person.

3.4.5 Evaluation

At this phase, we cross-validate between the original and our extracted age and gender of that particular person in the image and find the error rate.

4. RESULTS AND DISCUSSION

For the training of the model, we have been using the UTKFace dataset with expected predictions as,

Figure 2: UTKFace dataset sample

We expected the prediction to be accurate, according to the image shown. The error expected is to be minimum from the actual age. As each pixel is taken and features are extracted very carefully by the model, we get accurate results. The results after training the model and testing it with real-world data are,

Figure 3: 38 years old Rami Malik's picture

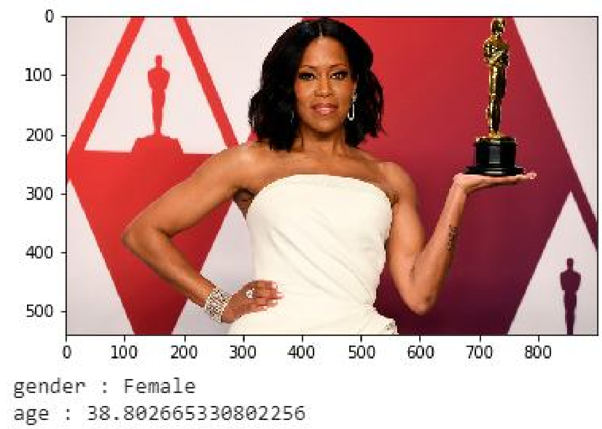

Figure 4: 49 years old Regina king's picture

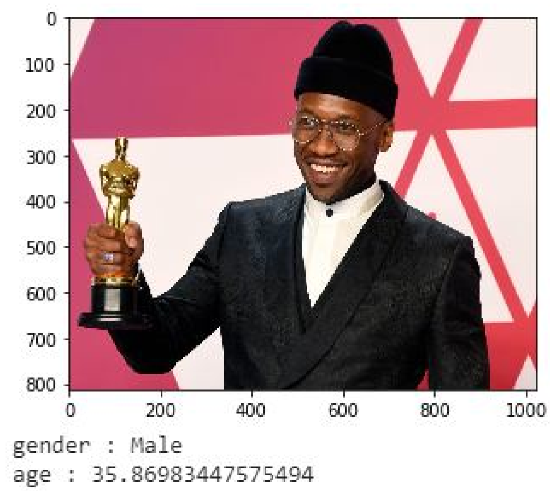

Figure 5: 46 years old Mahershala Ali's picture

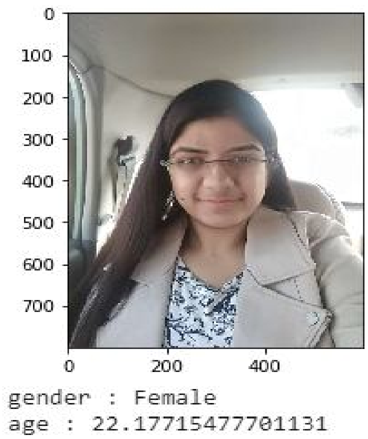

Figure 6: 22 years old First Author's picture

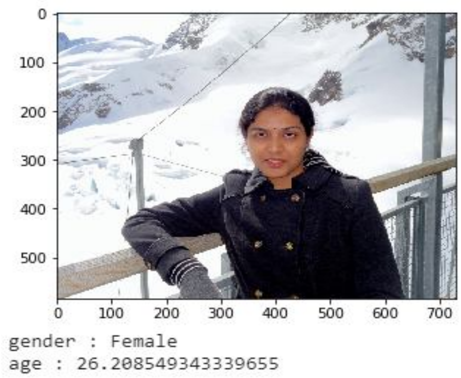

Figure 7: 25 years old female's high-quality, real-world picture

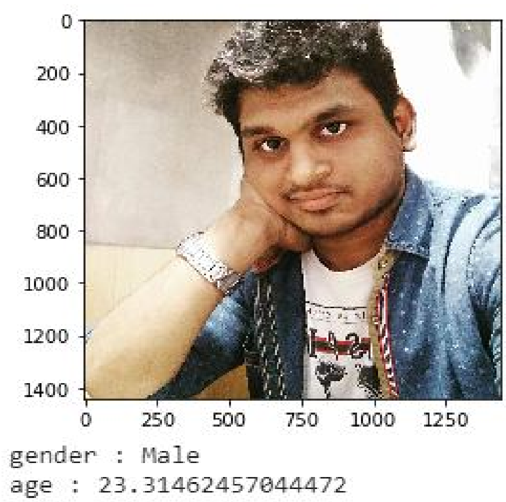

Figure 8: 22 years old male's low-quality real-world picture

The output clearly shows the ages of people from right to left. Their age and gender are predicted accurately by extracting the features from the image. The validation loss at each iteration in the training of the network is visualized in the form of a graph as below,

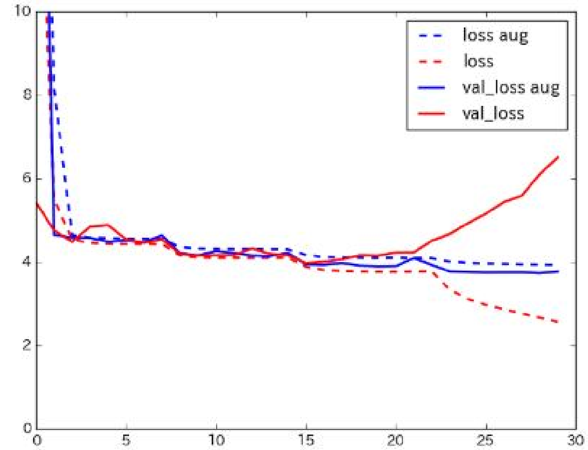

Figure 9: Loss Curve, epochs vs. the percentage

The x-axis represents the number of iterations, and the y-axis represents the loss at each corresponding iteration. Each loss is described in a different color [17] [18]. The best failure is observed from 3.5 to 4.2. We haven't found any overfitting for the fewer learning rates, that is, the iterations greater than 15.

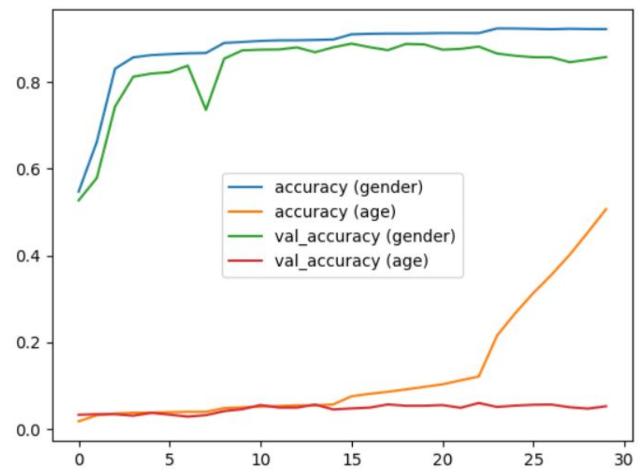

Figure 9: accuracy curve, epochs vs. percentage

Figure 9 represents the accuracy corresponding to age and gender. Age training accuracy increased drastically from 15th iteration, whereas validation accuracy stays put with a little fluctuation [19][20]. Within five repetitions, gender accuracy reached above 80% for both training and validation data. As gender training is a binary classification problem, the accuracy reaching to 80% is expected.

5. CONCLUSION

We have implemented an efficient method in the detection of gender and age. With the minimum NN, the identification process was easy, less time taking to compute, and consumed less memory compared to other methodologies in deep learning. The generated results are accurate and transparent with correct predictions; the gender of the person is correctly predicted a male or female. When we consider the age, due to the use of bias, the algorithm shows precise results. From the results in fig 4.7, we found that after the 20th iteration, the validation loss has increased, and the overall loss has decreased. In fig 4.8, we observe that the age accuracy increases after the 20th iteration. So, the validation loss is directly proportional to the age accuracy and indirectly proportional to the overall loss can be concluded from our experimentation

REFERENCES

1. Felix Anda, **Evaluating Automated Facial Age Estimation Techniques for Digital Forensics**, 2018 *IEEE Security and Privacy Workshops* https://doi.org/10.1109/SPW.2018.00028
2. M. H. Rahman, M. A. Bashar, F. H. M. Rafi, T. Rahman, and A. F. Mitul, "**An automatic face detection and gender identification from color images using logistic regression**," 2013 International Conference on Informatics, Electronics and Vision (ICIEV), Dhaka, 2013, pp. 1-6. https://doi.org/10.1109/ICIEV.2013.6572576
3. R.G.N.Meegama and J. C. Rajapakse, "**Curvature-based rational surface model for the detection of age and gender-related diversity in cortical morphology**," 3rd International Symposium on Image and Signal Processing and Analysis, 2003. ISPA 2003. Proceedings of the Rome, Italy, 2003, pp. 411-416 Vol.1.
4. C. Sarraute, P. Blanc, and J. Burroni, "**A study of age and gender seen through mobile phone usage patterns in Mexico,**"2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014), Beijing, 2014, pp. 836-843. https://doi.org/10.1109/ASONAM.2014.6921683
5. S. H. Lee, S. Hosseini, H. J. Kwon, J. Moon, H. I. Koo, and N. I. Cho, "**Age and gender estimation using deep residual learning network**," 2018 International Workshop on Advanced Image Technology (IWAIT), Chiang Mai, 2018, pp. 1-3.
6. S.Sudhakar, V.Vijayakumar, C.SathiyaKumar, V.Priya, LogeshRavi, V.Subramaniyaswamy, **Unmanned Aerial Vehicle (UAV) based Forest Fire Detection and monitoring for reducing false alarms in forest-fires**, Elsevier- Computer Communications 149 (2020) 1–16, https://doi.org/10.1016/j.comcom.2019.10.007
7. F.Becerra-Riera,H.Méndez-Vázquez,A.Morales-Gonzalez, and M. Tistarelli, "**Age and gender classification using local appearance descriptors from facial components,**" 2017 IEEE International Joint Conference on Biometrics (IJCB), Denver, CO, 2017, pp. 799-804. https://doi.org/10.1109/BTAS.2017.8272773
8. V.Cantoni, M.Porta, C.Galdi, M.Nappi, and H. Wechsler, "**Gender and Age Categorization Using Gaze Analysis**," 2014 Tenth International Conference on Signal-Image Technology and Internet-Based Systems, Marrakech, 2014, pp. 574-579. https://doi.org/10.1109/SITIS.2014.40
9. H.Sagha, J.Deng, and B.Schuller, "**The effect of personality trait, age, and gender on the performance of automatic speech valence recognition**," 2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII), San Antonio, TX, 2017, pp. 86-91.
10. C. Huang, Y. Chen, R. Lin, and C. -. J. Kuo, "**Age/gender classification with whole-component convolutional neural networks (WC-CNN),**" 2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Kuala Lumpur, 2017, pp. 1282-1285. https://doi.org/10.1109/APSIPA.2017.8282221
11. M. Uricár, R. Timofte, R. Rothe, J. Matas, and L. V. Gool, "**Structured Output SVM Prediction of Apparent Age, Gender and Smile from Deep Features**," 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Las Vegas, NV, 2016, pp. 730-738.
12. R.Vasanthi, R.Jayavadivel, K.Prasadh, J.Vellingiri, G.A kilarasu, S.Sudhakar, P.M.Balasubramaniam, **A novel user interaction middleware component system for ubiquitous soft computing environment by using the fuzzy agent computing system**, Journal of Ambient Intelligence and Humanized Computing (2020), Springer, doi.org/10.1007/s12652-020-01893-4.
13. Jagadeesh Gopal, Vellingiri J, Gitanjali J, Arivuselvan K, Sudhakar S, **An Improved Trusted On-Demand Multicast Routing with QoS for Wireless Networks**, International Journal of Advanced Trends in Computer Science and Engineering, Vol.9, No.1,January – February 2020,PP:261-265,https://doi.org/10.30534/ijatcse/2020/39912020.
14. Satheesh N, Sudha D, Suganthi D, Sudhakar S, Dhanaraj S, Sriram VP, Priya V,"**Certain improvements to Location aided packet marking and DDoS attacks in an internet**," Journal of Engineering Science and Technology Vol. 15, No. 1 (2020), pp: 94 - 107, School of Engineering, Taylor's University.
15. Sathiya Kumar C, Priya V, Sriram V P, Sankar Ganesh K, Murugan G, Devi Mani, Sudhakar S, "**An Efficient Algorithm for Quantum Key distribution with Secure Communication**, Journal of Engineering Science and Technology Vol. 15, No. 1 (2020), pp:77-93, School of Engineering, Taylor's University.
16. S.Sudhakar, N.Satheesh, S.Balu, Amireddy Srinish Reddy, G.Murugan, 2019, "**Optimizing Joins in a Map-Reduce for Data Storage and Retrieval Performance Analysis of Query Processing in HDFS for Big Data"**,International Journal of Advanced Trends in Computer Science and Engineering, (IJATCSE), Vol.8 No.5,pp:2062-2067,DOI:10.30534/ijatcse/2019/33852019.

17. K.B.,Prakash.**"Information extraction in current Indian web documents."** International Journal of and Technology(UAE), 2018: 68-71. https://doi.org/10.14419/ijet.v7i2.8.10332
18. Kolla B.P., Dorairangaswamy M.A., Rajaraman A. **"A neuron model for documents containing multilingual Indian texts."** 2010 International Conference on Computer and Communication Technology, ICCCT-2010,2010: 451-454. https://doi.org/10.1109/ICCCT.2010.5640489
19. Kolla B.P., Raman A.R. **"Data Engineered Content Extraction Studies for Indian Web Pages."** Advances in Intelligent Systems and Computing, 2019: 505-512. https://doi.org/10.1007/978-981-10-8055-5_45
20. Prakash K.B., Dorai Rangaswamy M.A.**"Content extraction studies using neural network and attributegeneration."** Indian Journal of Science and Technology, 2016: 1-10.
21. Prakash K.B., Dorai Rangaswamy M.A., Raman A.R. **"Text studies towards multi-lingual content mining for web communication."** Proceedings of the 2nd International Conference on Trendz in Information Sciences and Computing, TISC-2010, 2010: 28-31. https://doi.org/10.1109/TISC.2010.5714601
22. Prakash K.B.,Rangaswamy M.A.D. **"Content extraction of biological datasets using soft computing techniques.",** Journal of Medical Imaging and Health Informatics, 2016: 932-936. https://doi.org/10.1166/jmihi.2016.1931