ISSN 2278-3091

Volume 9, No.3, May - June 2020

International Journal of Advanced Trends in Computer Science and Engineering

Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse114932020.pdf

https://doi.org/10.30534/ijatcse/2020/114932020

Accurate Hand Gesture Recognition using CNN and RNN Approaches

Kolla Bhanu Prakash[1], Rama Krishna Eluri[2], Nalluri Brahma Naidu[3], Sri Hari Nallamala[4], Pragyaban Mishra[5], P Dharani[6]

[1,5,6]Department of Computer Science Engineering, K. L. Deemed to be University, India, drkbp@kluniversity.in pragnyaban@kluniversity.in, dharanipatibandla2004@gmail.com

[2]Department of Computer Science Engineering, Narasaraopeta Engineering College, India, rkeluri99@gmail.com

[3,4] Department of CSE, Vasireddy Venkatadri Institute of Technology, Guntur, Andhra Pradesh, nbnaidu1208@gmail.com, nallamala.srihari@gmail.com

ABSTRACT

Interaction between humans and computers, gesture recognition does play a critical role. By using modified convolutional neural network (CNN) and modified Recurrent Neural Networks (RNN) models for hand posture estimation, hand motion capture and hand object interaction in hand gesture recognition. By these models, we performed hand pose estimation on ICVL data set, motion capture available postures dataset and hand object interaction on kaggle dataset. RNN is for estimation and CNN is for feature extraction. A comparison of those three modules hand motion capture gives better accuracy.

Key words: Gesture Recognition, CNN, RNN, Motion Capture and Object interaction

1. INTRODUCTION

Gesture recognition is indeed one of the perceptual user interfaces that enable missions to capture and explain human movements. As it ordersit is intended to express a particular message for non-verbal communication [1]. Then the mission has the ability to understand the human gestures and execute specific code and for providing real-time data to a computer.

Hand gestures can sense tilting, rotation, and acceleration of movement by a controller that contains accelerometer and gyroscopes [2]. Human Hands are powerful devices in HCI applications. Recognition of the chore of hand movement has existed for many years and attracted many researchers in computer vision. The work in hand movement analysis includes hand pose estimate of hand posture, gesture recognition, hand motion capture and contact with hand object.

1.1 Hand poses estimation

In Human-computer Interaction the problem of estimation hand pose can always be done on a system for detecting the presence in particular to the practical significance. This interest is mainly grown with the most inexpensive range sensors and also brought the technology to a new level of performance.

Estimating hand pose problem can mainly be considered more difficult compared with estimating full body pose. That is because of more parameters and a more pronounced articulation. Self-occlusions and mutual occlusions are often wide and strong when communicating with various objects [30].

1.2 Hand motion capture

Hand motion capture is a well-defined research subject in applications industries and can capture the hand movement in a device with real-time capability.As cameras and images we can use an optical system to rack or capture motion or non-optical systems as inertial sensors for capturing or rotational data. From the perspective of performing motion capture in real time, calculation steps on graphics processing units are parallelised.

1.3 Hand object Interaction

Interaction of objects with analyzing of hands as reasoning how objects are being manipulated as separated and existing branches or research. In robotics, it finds direct applications and in the future, this proposal is enhancing the people interest and curiosity towards the automation with the development of wearable cameras and also analyzing the wild of self-interested video streams. Particular idea is more significant in the reconstruction of hand pose during interaction with objects from hand parts and also interested in one side from modeling.

2. RELATED WORK

Guangdong Hou, Runpeng Cui, Changshui Zhang, 2015 [3], proposed an approach for estimating the hand position with a recognized picture for the revival test. The pose estimation is designed for transfer Based on this algorithm, which takes into account a mixed criterion.

DongxuGao, ZhaojieJu, Jiangtao Cao, Honghai Liu, 2016 [4] studies make use of manipulation capability, grip consistency, And tightness of expressive hand with shapes from a picture, the human hand also serves as a fundamental component. Comprehensive approach has been introduced and is looking carefully to get the full view and appropriate details. Enough knowledge may be a guarantee of practical implementation in connection with human-machine (HCI), robot, and animation.

Henrique Weber, Claudio RositoJung, Dan Gelb, 2016 [5], Studies describe a novel processing and display technology that has inspired the creation of user interfaces that turn ordinary surfaces like walls and tables into immersive planes.The aim is to make computing resources available, customizable and allocated, and when the hand is laid flat, the color information and depth can effectively section the shape of the hand at varying distances between the hand and the surface.

Zhongxu Hu, Youmin Hu, BoWu, Jie Liu, 2017 [6] proposed on-the-spot study involves estimating RNN, CNN and neural gradient networks by using these techniques images with an end-to - end network, single frame length and contiguous frame sequence.Avoids composite Summary of images, feasible to input the actual image directly and avoids the complexity of data transformation during extracting features. The high assertions induced by a single-frame deformity, the nodes that are not connected between the layers.

Xingtai Fang, Xiaoyong Lei, 2017 [7] ,Studiesof the Hybrid CNN-AE Model Hand Pose Estimation based on a de-noising auto-encoder (AE) as a non-linear deep-learning embedding layer. Generative methods are enormous cost of computing optimization, and the fitness function is prone to an optimal solutions.

Ammar Ahmad, CyrilleMigniot, Albert Dipanda, 2017 [8], Given overview of hand object modelling with various applications to address the issue of hand gestures. They have been largely inspired by progress in monitoring and, in particular, in full human body detection, the estimation of the human body and the human hand has similarities. The exasperated burden of amplitude caused by increased power, the connectional specification on space of hand objects, causes self-occlusion due to the occurring manual poses and difficulty in hand rendering and hand object manipulation.

SatinderdeepKaur, NidhiBhatla, 2019 [9] ,Introduces a sign language, SIFT algorithm for retrieving hand features, and ABC – ANN is intended to increase the detection accuracy along with the collection and identification of category-based images.

KripeshAdhikari, Hamid Bouchachia, HammadiNait -Charif, 2019 [10] Proposed a different approach for the object tracking method that allows the human body to monitor the feedback from the machine learning algorithms to be implemented or explored during fall detection. Failed cases involve images not recovering in action recognition in the datasets, and body regions are not well specified, and overlapping.

3. METHODOLOGY

First, our actual framework is to represent hand gestures by using feature-computed manual shape and motion descriptors. The results of the evaluation show the appropriate initial of using data for performing gesture recognition system. Experiments are carried out on collected data, which contain a set of diverse perfect and rough gestures. In addition, the results of our approach prove helpful in the low bandwidth hand gesture recognition system, where a principal category is required.

Using a deep learning technique, we apply a transition teaching strategies to learn hand pose and system features from the depth image database generated for auto estimation. Second, we used recurrent deep learning techniques to estimate the spatial differences of the hand postures and their aspects.

Finally, precise prior detection information is merged with hand gesture recognition to perform precise detection in advance.

Dataset analyses provide the suggested technique is capable of detecting occurred gesture and of recognizing its form well before it ends.

3.1 Architecture

• Hand Pose Estimation: It is the taxonomy of existing deep learning alternative solutions and is typically unjustifiable in nature [11].

• This application needs an extra generative refinement to know the deep understanding of neural networks for the proposed system. Cutting edge frame works pipelines hand position involving deep learning with convolution feature extracting hand joint positions of depth map of hand.

Frame work represents flow from top to bottom and first step represent entire structure, here rough estimation was carried out with the joint parts having first level deep containers and multiple improvements with steps carried along with resolution containers by incorporating bottleneck hidden layers to solve the present limitations.

• Hand Object Interaction: examine the hand when they are interacting with the objects and also observing the objects how they are controlled is a interesting research topic[13].This focuses on robotics with clear advantages that have been intrigued by the growth of smartphone camera and growing demand for real-life automated processing of different video streams [14].

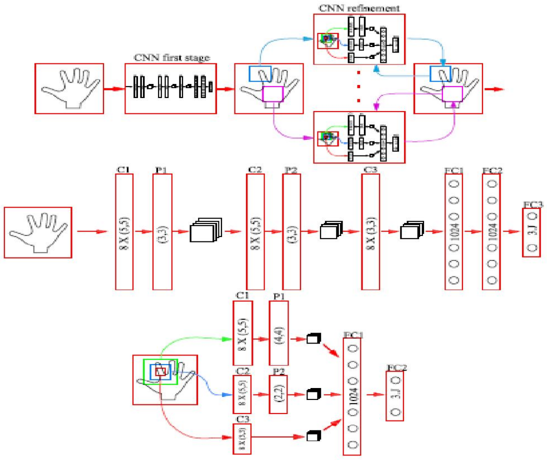

Figure 1: Hand Pose Assessment architecture by Oberweger et al. [12].

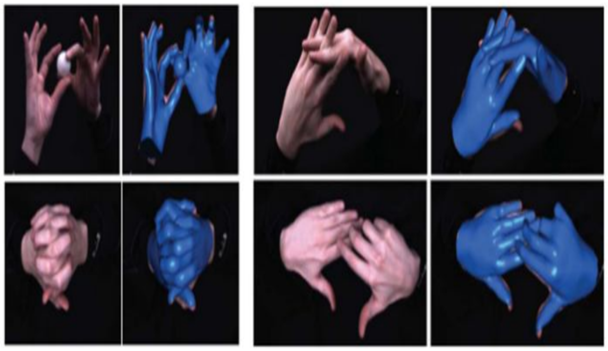

Figure 2: The pose of a hand occluded by an object and multi camera motion capture system.

This work is based primarily on modeling views. Analysis of occlusions for hand parts are done when communicating with objects which makes hand position rebuilding more important. And on other side these interactions with objects may introduce addition information for estimating the pose like any physical limitations. The Figure 1 gives the hand pose assessment architecture as specified in [12].The Figure 2 shows the pose of hand occluded by an object and multi camera motion capture system. Finally ,Figure 3 shows the hand skeleton into nine tuples of five joints representing hand structure which we compute SoCJ descriptor [15].

• Capture hand motion: Human tracking is done by using Machine Learning techniques in this movement capture and graphics are integrated. Complex learning problems are included in this process and less importance is given to collect large data sets for training. 2D and 3D model appearance based approach can be merged for fast and accurate estimation.

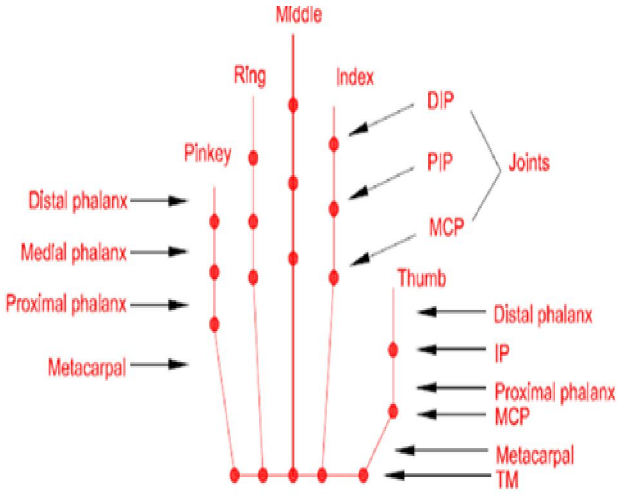

Figure 3: Hand skeleton into nine tuples of five joints representing hand structure which we compute SoCJ descriptor.[15]

3.2 Algorithm Description

Holistic hand pose estimation framework was proposed by Rosales [16] for conceive artificially synthetic data of hand motion multiple views using a Cyber Glove by capturing.

Algorithm 1: Hand Pose estimation

```
initialize votes Xi= Øfor joints i
for pixel ground(x, y, z) within image do
compute position of pixel (x, y, I (x, y)) →(a, b, c)
for every root no of tree then
Classify the root to attain the leaf l(x, y)
evaluation of distribution q(d| (x, y,z))
evaluate relevant body part d (i)
compute weight ω= q (d= d(i)).c^2
add set ofvote ((a, b, c),ω) set Yi
return for sample of votes Xi for joint i
```

This algorithm is used to predict the graded votes made it clear, before going through an alternative manner. Because of the root, the regression forest is used for continuous forecasts, and there is a regression forest at every node 1 used here. The root makes the predictions continuous.

To work out the joint positions directly Regression forest model is trained pixel-wise, in this every pixel of image (x, y, z)ЄS and corresponding leaf node 1 includes collection of similar votes xi Є Xl for all joint locations of i Є I and correct positions of pi from ground and all corresponding votes xi of all pixels in the training set (x, y, z)ЄS and all joint are integrated over i. To gain distribution voting globally for a joint given, probabilities of corresponding class pixels including depths as weighing coefficients are used to do the distribution invariant to the distance.

Algorithm 2: Hand Motion Capture

```
Compute   distribution of votes for every joint
for each joint i∈I do
Classify 2-components GMM for vote vectors
evaluate||μ1i−μ2i|| <τj do
Compute i-th joint is a high-determination
Compute the i-th  location of stronger component as mean
else
Then i-th joint is a low-determination
identify the nearest acquaintance of the high determination joints in G by Gaussian
The residual low-confidence joint locations from this Gaussian are to be updated
Then return Output
```

Fitting of into the distribution of every i-th vote and a GMM 2-component parameters _μi, Σ1i, ρ1i, μ2i, Σ2i, ρ2i|i∈ i_ in this the μ is means of corresponding values, Σ - variance and where ρ – weights of every component. And this takes us closer to defining the performance metrics, forecasts include both egalitarian and refinement-based models-and will be subject to next segment.

Algorithm 3: Hand Object Interaction

```
1: Y= Ø;
2: n= 0;
3: best feature selection
4: x1= argmax(I(Yn+ x));
5: Xn= Xn+ n+;
6: n= n+ 1;
7: Worst feature selection
8: x= argmaxx2Yn(I(Yn x));
9: Worst feature removing
10: if I(Yn x) >I(Yn) then
11: Yn+1 = Yn x;
12: n= n+ 1;
13: then got to7;
14: else
15: return3;
```

The above three steps detailed in algorithm that subset of Y along the score I are chosen to laid down automatically. We get a result of I of 75.73% with the use 10 SoCJs. The Figure 4 gives the flow chart for hand gesture recognition

3.3 Flow Chart

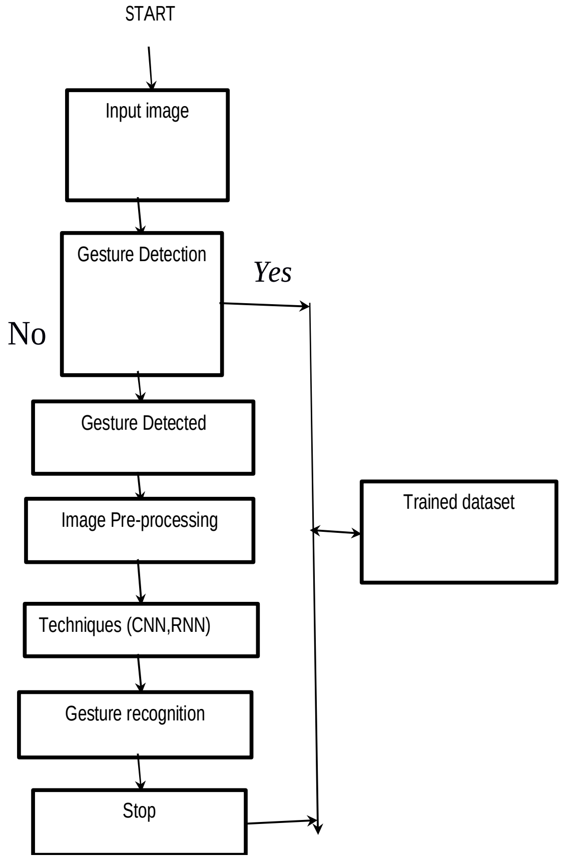

Figure 4: Flow Chart for Hand Gesture Recognition

4. EXPERIMENT RESULTS and ANALYSIS

Dataset: We evaluate our proposed methods on a "Motion Capture Hand Postures" dataset by OValery16 for hand motion capture. This dataset contains 38 attributes. X0, Y0, Z0, …, X11, Y11, Z11 are continuous attributes. User is a discrete feature that refers to the user ID. We only care about users with IDs 0, 1, 2, 5, 6, 8, 9, 10, 11, 12, 13, 14. That is a total of 12 users that we care about. Finally, Class is a discrete attribute with values 1, 2, 3, 4, 5 (5 values in total).

The ICVL Hand Pose Dataset contains 70,568 training frames, and 8,623 test frames [17]. The similarity measure covers 32 positions in the hand joint.

Multiple view CNNs contains 84,000 frames of 8 subjects for training and the remaining one dataset of 10,500 frames as test data collection [18].

We used a Kaggle Hand-Gesture Recognition Dataset. The dataset consists of 40,000 hand gestures captured by a Leap Motion sensor (size 620280). The photographs are divided into ten different subjects from ten different users, including five men and five women. The photos taken of right-hand movements are: L, thumb, side of the palm, finger, arm, C, OK and back [19].

We train and test our proposed RNN and CNN models on a machine with Intel Core i3 5930 K 3.50 GHz CPUs, 64 GB RAM, for Motion Capture Hand Postures. Create an RNN which takes the user information X0, Y0, Z0, ..., X4, Y4, Z4 as input attributes, returns the Class as output. Yet CNN just classifies Class 0**.**The classifier takes X0, Y0, Z0, ..., X4, Y4, Z4 as input attributes and returns the User ID as output. It can randomly divide the dataset into a training set (60%), a validation set (20%), and a test set (20%), and use some form of cross-validation [20].

CNN and RNN Models are trained and evaluated using the ICVL dataset to estimate 3D hand pose and these models are built within the PyTorch framework. The batch size of 16, the momentum as 0.9, and the weight decay as 0.0005 are chosen as network training parameters [21] .For 3D Regression network after 50 epochs 0.01 is set as learning rate divided by 10. And for the 3D U-Net, 0.001 is set as the learning limit.In order to prevent the training from stopping over fitting after 60 epochs [22] .Random initialisation was performed for all the weights of convolutionary layers in CNNs [23].

We evaluate and split of dataset commonly for the results in 32,000 images for training (60%), 4,000 for testing (20%), and 4,000 for validation (20%). Hyper parameters are used for training process for a similar implementation and down sampled a original image to (120X120) with batch size of 32.For fast and popular RMS Prop optimizer is use which observes well on all large datasets [24].

Result: Build a shallow NN-classifier that takes as input attributes X0, Y0, Z0, , X4, Y4,Z4, and returns the Class as the output. Inputs attributes X0, Y0, Z0, , X4, Y4, Z4 and the User information, and also returns the Class as the output. Shallow NN classifier ONLY for Class 0. The classifier takes as inputs attributes X0, Y0, Z0, , X4, Y4, Z4, and returns the User ID as the output [25].

The ICVL dataset results from our deep dense network with deep dense network based system [26] .The Table 1 gives the accuracy of motion picture for different epochs.We also compare the mean error distance of our method with those of the methods in as shown in Figure 5 and Figure 6, our method achieves the smallest mean error distance on most joints [27] The mean error distance over all joints of our method is 6.7 mm, while those of methods in are 9.3, 8.2, 7.3 and 6.8mm, receptively [28].

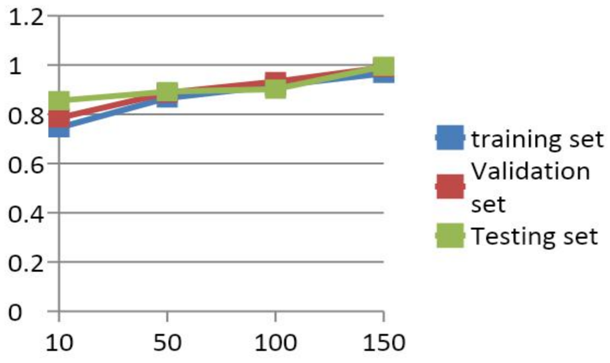

Figure 5: Comparison graph of accuracy of Training set, Validation Set and Testing Set

Table 1: Accuracy of motion capture hand postures dataset

Epoch	Training Set (accuracy)	Validation set(accuracy)	Testing set(accuracy)
10	0.7451	0.8451	0.8342
50	0.8961	0.8970	0.8810
100	0.9757	0.9770	0.9761
150	0.9658	0.9899	0.9906

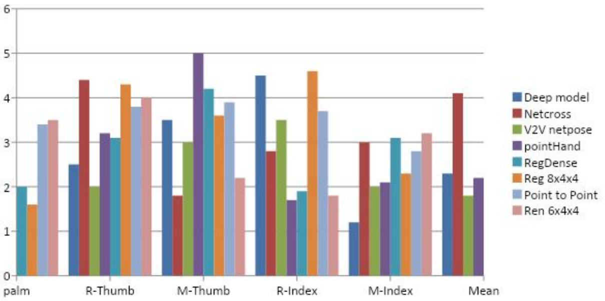

Figure 6: Comparison of ICVL dataset mean error (mm) and fraction of frames with in distance (%).

The test accuracy came out to be 91.2%, the top-2 accuracy was 98.7%, and the top-3 accuracy was 99.8% and out of 4,000 test images, 3,706 achieve images were correctly classified [29].

5. CONCLUSION

In this paper, we proposed the PyTorch method for manual pose estimation and hand motion capture in gesture recognition to be updated by CNN and RNN. The basic CNN layer here was the same as the extraction function, and later the extracted feature was input into the RNN model for estimation. Underpinned by a simple CNN network structure, it has proven that the RNN module can boost model accuracy through a series of interim experiments.

REFERENCES

1. https://www.marxentlabs.com/what-is-gesture-recognition-defined/
2. https://whatis.techtarget.com/definition/gesture-recognition
3. Guangdong Hou, Runpeng Cui, Changshui Zhang. "A Real-Time Hand Pose Estimation System with Retrieval", 2015 IEEE International Conference on Systems, Man, and Cybernetics, 2015.
4. DongxuGao, ZhaojieJu, Jiangtao Cao, Honghai Liu, “Towards Hand-Object Interaction Analysis From Depth Image “2016 Joint 8th International Conference on Soft Computing and Intelligent Systems and 2016 17th International Symposiumon Advanced Intelligent Systems.
5. Henrique Weber, Claudio Rosito Jung, Dan Gelb. "Hand and object segmentation from RGB-D images for interaction with planar surfaces", 2015 IEEE International Conference on Image Processing (ICIP), 2015.
6. Zhongxu Hu, Youmin Hu, Bo Wu, Jie Liu. "Hand Pose Estimation with CNN-RNN", 2017 European Conference on Electrical Engineering and Computer Science (EECS), 2017.
7. Xingtai Fang and Xiaoyong Lei,” Hand Pose Estimation on Hybrid CNN-AE Model,” International Conference on Information and Automation (ICIA) Macau SAR, China, July 2017.
8. Ammar Ahmad, CyrilleMigniot and Albert Dipanda, “Tracking Hands in Interaction with Objects: A Review “2017 13th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS).
9. Satinderdeep Kaur, Nidhi Bhatla, “An Efficient Gesture Recognition with ABC-ANN Classification and Key-Point Features Extraction for Hand Images “International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-10, August 2019.
10. Kripesh Adhikari, Hamid Bouchachia, Hammadi Nait-Charif, “ Deep Learning Based Fall Detection Using Simplified Human Posture,” World Academy of Science, Engineering and Technology International Journal of Computer and Systems Engineering Vol:13, No:5, 2019.
11. M. Oberweger, P. Wohlhart, and V. Lepetit, “Hands deep in deep learning for hand pose estimation, “CVWW, 2015.
12. M. Oberweger, P. Wohlhart, and V. Lepetit, “Training a feedback loop for hand pose estimation,”ICCV, 2015.
13. Natalia Neverova,” Deep Learning for Human Motion Analysis “Artificial Intelligence [cs.AI]. Université de Lyon, accessible in online http://theses.insa-lyon.fr/publication/2016LYSEI029/these.pdf2016.
14. L. Ballan, A. Taneja, J. Gall, L. van Gool, and M. Pollefeys, “Motion Capture of Hands in Action using Discriminative Salient Points”ECCV, 2012.
15. Quentin de Smedt, “Dynamic hand gesture recognition - From traditional handcrafted to recent deep learning approaches “,Computer Vision and Pattern Recognition [cs.CV],Université de Lille 1, Science set Technologies; accessible in online https://hal.archives-ouvertes.fr/tel-01691715CRIStAL UMR 9189, 2017.
16. R. Rosales, V. Athitsos, L. Sigal, and S. Sclaroff, “3D hand pose reconstruction using specialized mappings,” Computer Vision, 2000, 2001.
17. Prakash K.B., Rangaswamy M.A.D., Raman A.R., Statistical interpretation for mining hybrid regional web documents,2012,Communications in Computer and Information Science,292,CCIS,503-512.
18. Ismail M., Prakash K.B., Rao M.N.,Collaborative filtering-based recommendation of online social voting,2018,International Journal of Engineering and Technology(UAE),7,3,1504-1507.
19. Prakash K.B., Rajaraman A.,Mining of Bilingual Indian Web Documents,2016,Procedia Computer Science,89,514-520.
20. Prakash K.B.,Content extraction studies using total distance algorithm,2017,Proceedings of the 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology, iCATccT 2016,7912085,673-679.
21. Prakash, K.B., Mining issues in traditional indian web documents,2015,Indian Journal of Science and Technology,8(32),1-11.
22. Prakash, K.B., Dorai Rangaswamy, M.A., Ananthan, T.V., Rajavarman, V.N.,Information extraction in unstructured multilingual web documents,2015, Indian Journal of Science and Technology,8(16).
23. Prakash, K.B., Rangaswamy, M.A.D., Raja Raman, A.,ANN for multi-lingual regional web communication,2012,Lecture Notes in Computer

Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics),7667,LNCS,PART 5,pp-473-478.

24. Babitha, D., Ismail, M., Chowdhury, S., Govindaraj, R., & Prakash, K. B. (2020). Automated road safety surveillance system using hybrid cnn-lstm approach. International Journal of Advanced Trends in Computer Science and Engineering, 9(2), 1767-1773. doi:10.30534/ijatcse/2020/132922020
25. Babitha, D., Jayasankar, T., Sriram, V. P., Sudhakar, S., & Prakash, K. B. (2020). Speech emotion recognition using state-of-art learning algorithms. International Journal of Advanced Trends in Computer Science and Engineering, 9(2), 1340-1345. doi:10.30534/ijatcse/2020/67922020
26. Bharadwaj, Y. S. S., Rajaram, P., Sriram, V. P., Sudhakar, S., & Prakash, K. B. (2020). Effective handwritten digit recognition using deep convolution neural network. International Journal of Advanced Trends in Computer Science and Engineering, 9(2), 1335-1339. doi:10.30534/ijatcse/2020/66922020
27. Prakash, K., Lakshmi Kalyani, N., Vadla, P. K., & Naga Pawan, Y. V. R. (2020). Analysis of mammography for identifying cancer cells using convolution neural networks. International Journal of Advanced Trends in Computer Science and Engineering, 9(2), 1184-1188. doi:10.30534/ijatcse/2020/44922020
28. Reddy, A. V., Vege, H. K., & Prakash, K. B. (2020). Efficient and accurate hybrid deep learning model for multimodal disease risk prediction. International Journal of Advanced Trends in Computer Science and Engineering, 9(2), 1262-1267. doi:10.30534/ijatcse/2020/55922020
29. Vadla, P. K., & Prakash, K. B. (2020). Residue based adaptive resource provisioning through multi-criteria decision and horizontal scaling of vm’s in agent-based model for federated cloud. International Journal of Advanced Trends in Computer Science and Engineering, 9(2), 1610-1622. doi:10.30534/ijatcse/2020/108922020
30. Prakash, K. B., Dorai Rangaswamy, M. A., & Ananthan, T. V. (2014). Feature extraction studies in a heterogeneous web world. International Journal of Applied Engineering Research, 9(22), 16571-16579.