Atlantis Highlights in Computer Sciences, volume 4
Proceedings of the 3rd International Conference on Integrated Intelligent Computing Communication & Security (ICIIC 2021)

Automated Transcription for Raga Recognition and Classification in Indian Classical Music Using Machine Learning

Gowrishankar B S[1,*], Nagappa U Bhajantri[2]

[1] *Vidyavardhaka College of Engineering, Mysore, Karnataka, India*
[2] *Government College of Engineering, Chamarajnagar, Karnataka, India*
[*] *Corresponding author. Email: gowrish.vvce@gmail.com*

ABSTRACT

Raga recognition is only possible by trained musician to understand the notes based on the lead voice but a beginner is unable to decode the notes. This is significant for current scenarios in developing an automated note transcription system in Indian Classical Music (ICM). In the present research, various properties of raga and the machine learning techniques that are used for identifying the raga by a machine rather than a human or music expert are surveyed. The previously developed automatic raga recognition techniques using Carnatic and Hindustani Music, the main drawbacks and the improvements required are discussed. The present research work discusses about the future proposed models for automatic raga recognition using pitch detection algorithm, finding Tuning Offset, and Note Segmentation process. The proposed model will obtain better accuracy more than 96 % when compared to the existing CNN, GMM that obtained accuracy of 94 % and 95 %.

Keywords: *Automate, Automatic Raga recognition, Carnatic and Hindustani Music, Indian Classical Music, Machine learning techniques, Note transcription.*

1. INTRODUCTION

Raga is an important factor for Indian Classical Music (ICM) and it is the most significant element of melody from where new tones of music can be composed. ICM has two main branches known as Carnatic and Hindustani music which originated in the Indian subcontinent. Raga is the collection of melodic expressions that consist of identifiable melodic movements and the arrangement of tones that give rise to musical compositions and improvisations. The ragas are identified based on the levels of pitches and convey a particular kind of emotion [1, 2]. The arrangements of a set of pitches give rise to a set of ragas that establish a mood or flavor of the raga. Because there are no dependable guidelines in recognizing a raga, there are, however, generally, two methodologies by which individuals become acquainted with the raga recognition techniques. It ordinarily relies upon whether the individual is a prepared artist or the non-prepared yet educated individual. Individuals who have very little information on ragas can't recognize them except if they remember the synthesis and their ragas [3, 4]. Without comparing the rhythms generated from the previous swaras, it isn't possible for an individual for distinguishing a raga. There is a critical perception in this technique. Even though individuals can't communicate in a substantial way what a raga is, they however are capable of recognizing it. This certainty will lower the potential of a classifier due to sufficient information for each raga. The feature extraction process characterizes the set of parameters for different algorithms [5]. The classifier is used to relate the features obtained from feature extraction that classifies into respective notes to minimize the prediction error [6]. The audio signal is usually changing constantly with respect to time; each of the varied signals will have different amplitudes as the tone variations occur in the audio. The audio features are used for categorizing the low-level highlights and mid-level highlights. The low-level highlights are extricated by utilizing different sign preparing procedures while the mid-level highlights are removed on the low-level highlights [7, 8]. There were various researches undertaken for automatic raga recognition as it is helpful for many applications such as automated note transcription, recommendation of music, classification, etc.

There are various models where at least two ragas have somewhat very similar or a comparative arrangement of notes yet are completely different in the melodic impact they produce due to factors like the gamaka, worldly sequencing (which needs to submit to the limitations introduced in the arohana and avarohana), just as spots of Swara accentuation and rest [9-10]. Furthermore, raga distinguishing proof is a procured ability that requires critical preparation and practice. All things considered, programmed Raga order techniques have been broadly examined. In any case, existing strategies are dependent on Pitch Class that might ignore the data due to the wrong inclination that is discussed in section 2.

The structure of the paper is as follows. Section 2 describes the existing methodologies involved for raga recognition in Carnatic music using machine learning algorithms. Section 3 describes the problems that occurred in the existing model and the solutions for the model are described in section 4. Section 5 describes the conclusion and future works.

2. RELATED WORKS

The given raga has a melodic phrase and each raga was manipulated with Gamakas that were not identified in the existing models. For example, a single raga can be used for tuning more numbers compositions by different people. Each tone has its melody but the base of the composed tone will be from one single raga which was used as a reference for composition. Similarly, the set of notes present in each raga can be the same but it sounds different based on the properties of every swara present. Each swara has its sound or melody that varies in each of the ragas. The various properties of ragas include the order of notes which is known as the ascending or descending pattern also known as arohana or avarohana. The second property includes gamakas which involve the variation of the pitch of a note, using heavy forceful oscillations between adjacent and distant notes. The gamaks have their sort of relative strength where each of the Swara positions has its duration. Figure 1 shows the flow chart for the general raga recognition procedure.

2.1. Arohana and Avarohana: The Ascending and Descending Progressions of a Raga

The recognition of raga is done based on the ascending and the descending order of notes progressions which in musical terms are known as arohana and avarohana. Certain rules need to be followed precisely and to the point before singing the notes as each of them may look the same in the writing but may have a different sound. This is known as a transition between notes and it is necessary to follow these transitions before reciting. The transitions generally occur with a note that is near to the current note in arohana/avarohana.

2.2. Gamakas

Gamakas are having a fixed frequency value that operates with a rapid oscillatory movement for each note. These notes have their pitch tone movements and together form gamakas. The gamakas have another form that involves in making the sliding movement of one note to other notes.

2.3. Various Roles Played by the Notes

The swaras present in a raga do not play the same role throughout the music. The ragas have the same constituent notes but the functionality of the constituent swara varies that altogether leads to a different feeling and rhythm of a raga. The properties enhance the character of tones and it takes the shape for setting the mood depending on the Swara constituents, which is shown in table 1.

Table 1. Properties of Raga

Definition	Generic names of ragas
Swaras bring out the mood of the raga.	Jeeva swaras
Notes which appear at the end of musical phrase	Nyasa swaras
The note which occurs at the beginning of the melodic phrases	Graha swara
Swara that occurs relatively frequently	Hamsa swara
Prolonged notes.	Dirgha swaras

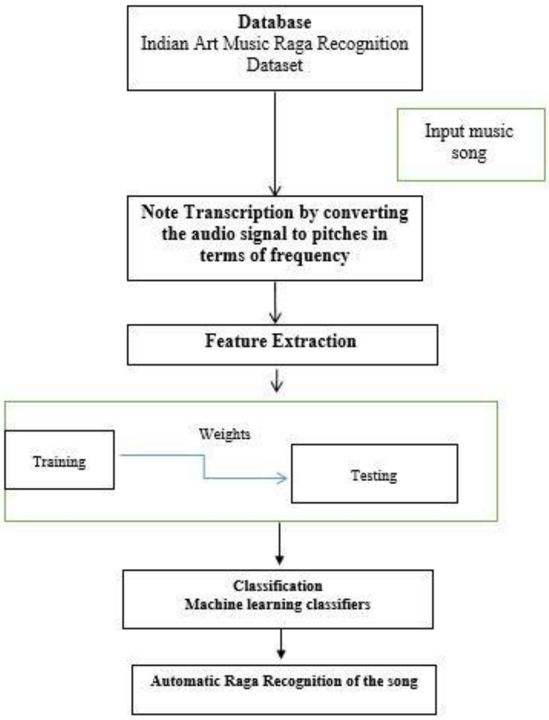

Figure 1. General flow diagram for raga recognition

Gamakas and pitch extraction appropriately accurately represents the Gamakas. The problems faced in the classification of ragas were difficult to handle due to the Gamakas of a note. Therefore, a pitch extraction system will improve the identification of the Gamakas when they are considered as features. In [4], Automatic classification of Carnatic music instruments was developed using Linear Predictive Coding (LPC) and Mel Frequency Cepstral Coefficient (MFCC). The well-known features were extracted and analyzed the capability of the features from different instruments. The classification techniques like, Artificial Neural Network, Support Vector Machine, and Bayesian classifiers were fed for features.

Skipping tonic detection

The tonic identification sometimes creates an error that would affect the performance of the raga recognition system. Therefore, eliminating such absolute tonic frequency is important for raga recognition. In [5], Bhavya Mor modeled Indian music using a variable order gaped Hidden Markov Model (HMM) removal of the base frequency of an instrument or a singer is necessary for the identification of raga. However, the model when applied to a larger musical sequence database it increased the maximum length of sub-sequences with larger dependencies.

Resolution of pitch-classes

Gamakas are the microtonal variations that play an important role in the perception of ICM and the transition involved in the gamaka as well as the note is challenging to determine. The information should be obtained partially at least for determining the pitch distributions. Anu S [6] developed a model for the transformation of architectural forms into music and vice-versa for the enhancement of the musical and architectural libraries. However, the logic was used in the research used in the development related software still required enhancement of substructure libraries, which would help architects and music composers.

Sarkar *et al.* [7] developed raga identification from Hindustani Classical Music signal based on the compositional properties. In the existing methods, identification of raga automatically utilized more time and therefore, to overcome such an issue, the developed model introduced a co-occurrence matrix for summarizing the problem. The features from the audio clip from the ICM dataset were considered for identifying properties of raga features The features were fed into the Support Vector Machine (SVM) classifier for classifying the type of raga the audio clip came under. However, the classification errors had occurred as manual raga identification required expertise suggestions which were not easily available.

Makarand Velankar *et al.* [8] performed pattern recognition of the tone melodies in ICM for raga identification. From the analysis done on the existing methods, the researchers have taken initiative for identifying raga with the least sample duration. In the developed model, the existing problem was overcome which resulted in less computational time and cost. The developed model extracted the pitch-based features for identifying the melodic level by using the auto correlation method. The results showed that the developed models required varied duration of time for accurate identification. However, the results were not enough for identifying the raga accurately as a limited number of raga samples were utilized for the research.

Sanchali Das *et al.* [9] developed a computational model that grouped the ragas based on the moods using machine learning techniques. The proposed technique used features such as Rhythm, Intensity, and Timbre to group the classes, but the tones showed contrast with other Indian tones and western tones. The developed model used audio features from each of the songs analyzed mathematically which gave better results having similar range into separate subclasses. However, the ground truth values for amplitude-based features limited the performance of the model.

Medina *et al.* [10] developed an emotional characterization of music utilizing neural organizations with the Medieval dataset. The current models utilized Multilayer Perceptron (MLP) which was prepared with the unreservedly accessible Medieval information base that was insufficient for procuring a decent arrangement in the results as the upsides of valence and excitement imbalanced order. However, the characteristics of the dataset-required size, class balance, quality of the annotations were still more needed to improve for achieving good performance in terms of accuracy.

Kaur and Kumar [11] developed Mean Centered Clustering (MCC) that was utilized for tune grouping for automated raga recognition. The created MCC procedure boosted the distance among the tones and decreased the spread of information in singular groups. The created research included pre-processing methods and classifiers Artificial Neural Network (ANN) and SVM acquired better outcomes for tune grouping. However, the identification or extraction of appropriate musical sample features was further needed to enhance the success rate of the classifier.

John *et al.* [12] performed classification of ICM based data on automatic Raga recognition using a Deep Learning model through audio signal processing. The model classified the music signal that managed the audio dataset for music therapy. The research work utilized Convolution Neural Network (CNN) for extracting the pitch contour features for the selection of ragas from Carnatic music. However, the developed model faced challenges during pattern recognition among the obtained pitches as the sequence of pitches was based on the Parsel-mouth library.

Manjabhat *et al.* [13] developed a Feed-Forward Neural Network organized model with Gaussian Mixture Models (GMM) and Choice Trees for distinguishing raga and tone. The developed numerical model was based on the pitch boundaries that distinguished the tones for raga recognition and were freely resolved based on the extracted pitches using the histogram technique. The Probability Density Function (PDF) is used for pitch estimation which separates the extreme level pitches from the music cut. By utilizing various classifiers, the raga was distinguished but still, the precision was lower as more information was lost during arranging the notes resulted in over-fitting issues.

Sinith *et al.* [14] presented the Fibonacci arrangement-based pitch dispersion Hidden Markov Model. The pitch shapes were dependent on a table which was inferred by utilizing the Fibonacci arrangement for ICM. These pitch forms adequately addressed the song for music data recovery. The ragas consisted of more regular swaras that were distinguished in a wrong manner and also the ragas having more than two distinctive swaras were neglected, which resulted in lesser accuracy.

3. METHODS

Initially the dataset has to consider the input data from Comp Music Dataset [15]. The dataset comprises of full-length audio recordings having their respective raga labels and are useful for evaluating the approaches for automatic raga recognition in Indian art music. In this research, CMD is available in the Comp Music group which consists of high-quality audio recordings taken from ICM concerts [16-19]. The dataset consists of 124 hours of time duration that includes 480 audio recordings, which are available in mp3 format. The 12 full lengths of each raga recording are present in the dataset, which is having 40 types of distinct ragas [20]. Here, diverse refers to the melodic attributes, therefore, training and evaluation of the model using the dataset give better results. HMD comprises 116 hours of audio recordings and editorial metadata [21], which is presented among 300 recordings having 30 types of ragas [22].

3.1. Pitch Extraction

The pitch detection algorithm needs to be designed predominately for the sampled inputted audio file. Using the algorithm, the pitch finder tracks the melodic voice precisely and performs pitch adjustments. To develop a symphonic example to coordinate with dynamic programming based on a smoothing strategy [23].

3.2. Finding the Tuning Offset

To convert all the pitch values folding to a single octave using finely-binned histogram maximum provides us the underlying tuning offset of the audio. The main importance of tuning offset is to normalize the continuous pitch contour for tuning the standard frequency based on the simple vertical shift without any quantization to the note grid at this point [24-27].

3.3. Note Segmentation

The continuous oscillations that are marked by the pitch contour are connected with more stable pitch regions. The melodic ornamentation in ICM is diverse and elaborate investigates the notes to relatively make stable among the varying pitch for detecting stable note regions that are having varying pitch contour [28].

3.4. Distance Measure

To perform accurately the rhythm revolution from one distribution to other, it accomplishes an arrangement for tonic note span with that of the other tone [29]. Since data about the tonic note of each tune is not accessible, all potential arrangements between two pitch class profiles should be considered and the one that matches best as far is defined with the distance measure.

4. RESULTS

Table 2 shows the comparative analysis of the existing models describing the methodology, dataset, and the performances achieved by these models.

Table 2. Comparative analysis of the existing models

Authors	Dataset	Methodology	Accuracy	Error Values	Standard Deviation	F-measure
Velankar [7]	CompMusic dataset	N-gram approaches	91.66	8.34	-	-
Sanchali Das [8]		Statistical modeling	49 to 63%.	-	0.037	-
Yesid Ospitia-Medina [9]	MediaEval	Multilayer Perceptron	75%	-	-	73

John [12]	CompMusic dataset	CNN	94%	6	-	-
Samsekai Manjabhat [13]		GMM	95 %	5	-	-
Sinith [14]		Fibonacci series based pitch distribution HMM	95.3%	4.7	-	-

5. DISCUSSION

The results in [7] showed the model required varied duration of time for better accuracy but the results were not enough for identifying the raga accurately. This was due to less number of raga samples utilized in the research and obtained accuracy of 91.66 % and Error value of 8.34. The developed model results obtained by [8] utilized various audio features were mathematically analyzed considered the ground truth values. But for amplitude-based features limited the performance of the model and obtained accuracy ranging from 49 to 63% and Standard deviation of 0.037 The developed model results from [9] were obtained based on characteristics of the dataset-required size, class balance, quality of the annotations required improvement for achieving good performance in terms of accuracy. The results obtained for the model in terms of accuracy was of 75% and F-measure of 73 % In [11] developed model obtained results based on the pattern recognition among the obtained pitches an accuracy 94% and Error value of 6. In [12] developed a model that utilized various classifiers for raga recognition distinguished lowered more information and resulted in over-fitting issues. Thus, the results obtained in terms of accuracy was of 95 % and Error value of 5. The developed model results obtained in [13] considered more regular swaras distinguished results and also the ragas having more than two distinctive swaras were neglected, which resulted in lesser accuracy of 95.3% and error value of 4.7. Whereas, the proposed model performs automatic raga recognition using pitch detection algorithm, finding Tuning Offset, and Note Segmentation processes achieves accuracy more than 96 %.

6. CONCLUSION

The present research work presented comprehensive concepts about ragas and their properties. The existing raga recognition techniques are surveyed under which their contributions and their approaches are being focused upon. The main key aspect is to address and outline the methodologies involved and discussed. The contributions of the works are highlighted with details such as dataset testing that provides the processing steps for pitch extraction procedure in Carnatic music. The present research work discussed the pre-processing techniques involved in pitch extraction using Carnatic music. The dataset utilized for the existing researches was of the polyphonic audio recordings for raga recognition. The automatic transitions of gamakas were discarded and were utilized as an additional feature for raga recognition. The present research work discusses about the future proposed models for automatic raga recognition using pitch detection algorithm, finding Tuning Offset, and Note Segmentation process. The proposed model will obtain better accuracy more than 96 % when compared to the existing CNN, GMM that obtained accuracy of 94 % and 95 %. Further inclusion of features for characteristic phrases and gamakas recognition suggests that the temporal features are important for the future.

REFERENCES

[1] A.K. Patel, S.S. Chouhan, R. Niyogi, Using Crowd-Sourced Data for Music Mood Classification, in: A. Mondal, H. Gupta, J. Srivastava, P. Reddy and D. Somayajulu, Proceedings of the 6th International Conference on Big Data Analytics, Lecture Notes in Computer Science, vol. 11297, 2018, pp. 363–375. https://doi.org/10.1007/978-3-030-04780-1_25

[2] B.G. Patra, D. Das, S. Bandyopadhyay, Labeling data and developing a supervised framework for Hindi music mood analysis, Journal of Intelligent Information Systems, vol. 48, no. 3, 2017, pp. 633–651. https://doi.org/10.1007/s10844-016-0436-1

[3] R. Sarno, J.A. Ridoean, D. Sunaryono, D.R. Wijaya, Classification of music mood using MPEG-7 audio features and SVM with confidence interval, International Journal on Artificial Intelligence Tools, vol. 27, no. 05, 2018, pp. 1850016. https://doi.org/10.1142/S0218213018500161

[4] S. Shetty, and S. Hegde, Automatic classification of carnatic music instruments using MFCC and LPC. In Data Management, Analytics and Innovation pp. 463-474, 2020. https://doi.org/10.1007/978-981-32-9949-8_32

[5] B. Mor, S. Garhwal, A. and Kumar, MIMVOGUE: modeling Indian music using a variable order gapped HMM. Multimedia Tools and Applications, vol. 80, no. 10, pp.14853-14866, 2021. https://doi.org/10.1007/s11042-020-10303-y

[6] S. Anu, K. Muthukkumaran, M. Punniyamoorthy, S.A. Veerapandian, and G. Sangeetha, A methodology for the transformation of architectural

forms into music and vice-versa for the enhancement of the musical and architectural libraries. Multimedia Tools and Applications, vol. 80, no. 7, pp. 10901-10926, 2021. https://doi.org/10.1007/s11042-020-10201-3

[7] R. Sarkar, S.K. Naskar, S.K. Saha, Raga identification from Hindustani classical music signal using compositional properties, Computing and Visualization in Science, vol. 22, no. 1-4, 2019, pp. 15–26. https://doi.org/10.1007/s00791-017-0282-x

[8] M. Velankar, A. Deshpande, P. Kulkarni, Melodic pattern recognition in Indian classical music for raga identification, International Journal of Information Technology, vol. 13, 2018, pp. 251–258. https://doi.org/10.1007/s41870-018-0245-6

[9] S. Das, B.K. Bhattacharyya, S. Debbarma, Building a computational model for mood classification of music by integrating an asymptotic approach with the machine learning techniques, vol. 12, 2020, pp. 5955–5967. https://doi.org/10.1007/s12652-020-02145-1

[10] Y.O. Medina, J.R. Beltrán, S. Baldassarri, Emotional classification of music using neural networks with the MediaEval dataset, Personal and Ubiquitous Computing, 2020, pp. 1–13. https://doi.org/10.1007/s00779-020-01393-4

[11] C. Kaur, R. Kumar, Mean centred clustering: improving melody classification using time-and frequency-domain supervised clustering, Sādhanā, vol. 44, no. 2, 2019, pp. 40. https://doi.org/10.1007/s12046-018-1001-0

[12] S. John, M.S. Sinith, R.S. Sudheesh, P.P. Lalu, Classification of Indian Classical Carnatic Music Based on Raga Using Deep Learning, in: Proceedings of thae 5th International Conference on Recent Advances in Intelligent Computational Systems (RAICS), IEEE, 2020, pp. 110–113. DOI: 10.1109/RAICS51191.2020.9332482

[13] S.S. Manjabhat, S.G. Koolagudi, K.S. Rao, P.B. Ramteke, Raga and tonic identification in carnatic music, Journal of New Music Research, vol. 46, no. 3, 2017, pp. 229–245. https://doi.org/10.1080/09298215.2017.1330351

[14] M.S. Sinith, S. Tripathi, K.V.V. Murthy, Raga recognition using fibonacci series based pitch distribution in Indian Classical Music, Applied Acoustics, vol. 167, 2020, pp. 107381.

[15] S. Gulati, J. Serrà, K.K. Ganguli, S. Sentürk, X. Serra, Time-delayed melody surfaces for raga recognition. In Proceedings of the 17th International Society for Music Information Retrieval Conference (ISMIR), pp. 751–757, 2021.

[16] Z． Guo, K. Yu, Y. Li, G. Srivastava, and J. C. -W. Lin, “Deep Learning-Embedded Social Internet of Things for Ambiguity-Aware Social Recommendations”, IEEE Transactions on Network Science and Engineering, doi: 10.1109/TNSE.2021.3049262.

[17] K. Yu, L. Tan, M. Aloqaily, H. Yang, and Y. Jararweh, “Blockchain-Enhanced Data Sharing with Traceable and Direct Revocation in IIoT”, IEEE Transactions on Industrial Informatics, doi: 10.1109/TII.2021.3049141.

[18] K. Yu, L. Lin, M. Alazab, L. Tan, B. Gu, “Deep Learning-Based Traffic Safety Solution for a Mixture of Autonomous and Manual Vehicles in a 5G-Enabled Intelligent Transportation System”, IEEE Transactions on Intelligent Transportation Systems, doi: 10.1109/TITS.2020.3042504.

[19] Parameshachari, B. D., H. T. Panduranga, and Silvia liberata Ullo. "Analysis and computation of encryption technique to enhance security of medical images." In *IOP Conference Series: Materials Science and Engineering*, vol. 925, no. 1, p. 012028. IOP Publishing, 2020.

[20] Rajendran, Ganesh B., Uma M. Kumarasamy, Chiara Zarro, Parameshachari B. Divakarachari, and Silvia L. Ullo. "Land-use and land-cover classification using a human group-based particle swarm optimization algorithm with an LSTM Classifier on hybrid pre-processing remote-sensing images." *Remote Sensing* 12, no. 24 (2020): 4135.

[21] Arun, M., E. Baraneetharan, A. Kanchana, and S. Prabu. "Detection and monitoring of the asymptotic COVID-19 patients using IoT devices and sensors." *International Journal of Pervasive Computing and Communications* (2020).

[22] L. Zhen, Y. Zhang, K. Yu, N. Kumar, A. Barnawi and Y. Xie, "Early Collision Detection for Massive Random Access in Satellite-Based Internet of Things," IEEE Transactions on Vehicular Technology, vol. 70, no. 5, pp. 5184-5189, May 2021, doi: 10.1109/TVT.2021.3076015.

[23] L. Tan, K. Yu, A. K. Bashir, X. Cheng, F. Ming, L. Zhao, X. Zhou, “Towards Real-time and Efficient Cardiovascular Monitoring for COVID-19 Patients by 5G-Enabled Wearable Medical Devices: A Deep Learning Approach”, Neural Computing and Applications, 2021, https://doi.org/10.1007/s00521-021-06219-9

[24] Kumar, M. Keerthi, B. D. Parameshachari, S. Prabu, and Silvia liberata Ullo. "Comparative Analysis to Identify Efficient Technique for Interfacing BCI System." In *IOP Conference Series: Materials Science and Engineering*, vol. 925, no. 1, p. 012062. IOP Publishing, 2020.

[25] Subramani, Prabu, Ganesh Babu Rajendran, Jewel Sengupta, Rocío Pérez de Prado, and Parameshachari Bidare Divakarachari. "A block bi-diagonalization-based pre-coding for indoor multiple-input-multiple-output-visible light communication system." *Energies* 13, no. 13 (2020): 3466.

[26] Nguyen, Tu N., Bing-Hong Liu, Nam P. Nguyen, and Jung-Te Chou. "Cyber security of smart grid: attacks and defenses." In *ICC 2020-2020 IEEE International Conference on Communications (ICC)*, pp. 1-6. IEEE, 2020.

[27] K. Yu, M. Arifuzzaman, Z. Wen, D. Zhang and T. Sato, “A Key Management Scheme for Secure Communications of Information Centric Advanced Metering Infrastructure in Smart Grid,” IEEE Transactions on Instrumentation and Measurement, vol. 64, no. 8, pp. 2072-2085, August 2015. https://ieeexplore.ieee.org/document/7138617

[28] Pham, Dung V., Giang L. Nguyen, Tu N. Nguyen, Canh V. Pham, and Anh V. Nguyen. "Multi-topic misinformation blocking with budget constraint on online social networks." *IEEE Access* 8 (2020): 78879-78889.

[29] Naeem, Muhammad Ali, Tu N. Nguyen, Rashid Ali, Korhan Cengiz, Yahui Meng, and Tahir Khurshaid. "Hybrid Cache Management in IoT-based Named Data Networking." *IEEE Internet of Things Journal* (2021).