Cover Song Identification through Symbolic Representation and Classifier

D. Khasim Vali, Nagappa U. Bhajantri

Abstract: In this work, we propose a work erected to cover song identification riding through a Symbolic Classifier. The classification of a cover song, which is a different variety of an earlier noted song for music retrieval, has received extra consideration. However, there is an urgency to protect the interest of melody music contributors. In view of this, several efforts have been heaping in literature. Systems for Classifying a cover song typically involved in Preprocessing, Extraction of Chroma features and Finally a Symbolic classifier for Identification.

Keywords: Beat Tracking, Chroma Features, Classifiers, Symbolic Representation.

I. INTRODUCTION

A spread tune, or essentially spread, is another form of standing music recorded or organized by another performer. A spread reuses the tune and verses of the first melody, however it is performed with new vocalists and instruments. Further melodic issues, for example, key, mood, and kind can be reinterpreted by the new craftsman. Since the copyright of piece of the spread still has a place with the writer of the first melody, discharging a spread tune without authorization of the real supporter may cause a legitimate clash.

Another case is music testing, which is the demonstration of procedure that reuses a bit of existing sound accounts. The inspecting is broadly viewed as a method for recovering music today, yet permitting that the first maker approves its reuse is a lawful prerequisite. At the end of the day, Cover melody distinguishing proof is an errand that intends to measure the comparability between two tunes. It very well may be utilized to counteract the encroachment of copyright, and furthermore to be a target reference if there should arise an occurrence of contention. For decade numerous methodologies for spread tune distinguishing proof have been proposed [9, 17, 19]. People for the most part perceive the spread through the melodic or verse likeness, however partition of the prevalent tune from a blended music sign is as yet not at a dependable level, and extraction of the verses can be endeavored just on the off chance that it is plainly isolated.

* Correspondence Author
   D. Khasim Vali*, Department of computer Science and Engineering, Vidyavardhaka College of Engineering, Mysuru, India,
     E-mail: dkv@vvce.ac.in.
   Nagappa U. Bhajantri, Department of computer Science and Engineering, Government College of Engineering, Chamarajanagara, India,
     E-mail: bhajan3nu@gmail.com.

The created dataset "covers30", contains 30 groups of original and cover songs - crossing across styles, live, genres and documented music, it is biased towards regional languages. Most of songs contain a cover version however some of songs contain up to three.

Similarly, the extended "covers80" [14] employed at MIREX 2007 to benchmark cover song recognition systems. The dataset contains 80 groups of original and cover songs total in 166 which comprises styles, live, genres and documented music. In other words covers80 predominantly, oriented towards western music.

Rest of the work is prepared in the subsequent way. In section II we present a deep coverage on broader area of cover song identification. The emphasize of proposal strategy is in section III. Section IV dedicated to comment on Chroma features extraction. Further, Symbolic Representation and Identification in section V. Outcomes are quantified with a yardstick via empirically as well as statistically narrated and experimental validations are offered in section VI. Conclusions are drawn in section VII.

II. RELATED WORK

Over the last years, in the space of cover song identification, there has been a considerable amount of new approaches and techniques that try to handle different issues [2, 4, 14]. The typical goal is to try new algorithms or combinations of them in order to increase the results in comparison to previous systems, but the recent main focus by most researchers has been towards scalable strategies [21, 22]. The most common way to calculate the similarity between two different songs is through the use of alignment-based methods and they have shown to produce good results, 75% MAP in MIREX'2009[1].

However, these methods are computational expensive and, when applied to large databases, they can become impractical: the best performing algorithm [1] in MIREX'2008 implemented a modified version of the Smith-Waterman algorithm and taken approximately 104 hours to compute the results for 1,000 songs. If algorithm applied to the Million Song Dataset (MSD), the estimated time to conclude would be of 6 years [2]. Martin et al. [3] suggest the use of Basic Local Alignment Search Tool (BLAST), a bioinformatics sequence searching algorithm, as an alternative to dynamic programming solutions.

The data is indexed based in similarity between songs, and to compute the similarity value, the best subsequences are chosen, and then compared. Khadkevich et al. [4] extract information about chords and store them using Locality-Sensitive Hashing (LSH). Bertin-Mahieux et al. [3] adopt the 2D Fourier Transform Magnitude for large-scale cover detection. This solution was further improved by Humphrey et al. [5], who modified the original work to use a sparse, high-dimensional data-driven component, and a supervised reduction of dimensions.

The authors J.V. Balen et al. [2] extract high-level musical features that describe harmony, melody, and rhythm of a musical piece. Those descriptors are stored with LSH, which allows retrieving the most similar songs. Lu and Cabrera [6] use hierarchical K-means clustering on Chroma features to find audio words, in other words centroids. A song will then be represented by its audio words. Moreover similarity with other songs will be determined by audio words share with the same location.

Outside the field of large-scale cover identification, several solutions regarding distance fusion [4] have been suggested. Salamon et al. [7] extract the melodic line, the bass-line, and HPCP 12-bins for each song. They explore the fusion of those features in order to discover which results in the best performance. Distance fusion is also the main focus in the effort of Degani et al. [6], where they recommend a heuristic for distance fusion. Their proposal consists of normalizing all values to [0, 1], computing a refined distance value, and produce a single matrix of results.

The customary methodologies depicted above compute the separation between an inquiry and the tunes to be looked at, and verify that the melody with the closest separation is almost certain to be a spread. Since this procedure is independent from each question, the outcome from "another rendition of a similar spread" can't be considered. On the off chance that it is potential, tunes with various lengths can be spoken to in a similar space.

III. PROPOSED METHOD

The planned system, shown in Figure 1, portrayed in three stages. Firstly, the preprocessing stage, attempts to convert audio signals into chroma features for each song. Then reduce the chroma features through DCT, FFT and DWT coefficients. Subsequently diminished features will fed into various classifiers such as SVM, NB and KNN.

A. Beat Tracking

Since raised methodology essential portrayal comprises of a component vector for each beat, it must beginning by distinguishing the beat division times in the music sound. The principal phase of beat following proselytes the sound into a "beginning quality" esteem at a 250 Hz inspecting rate. Then again, this is determined by taking the primary request distinction alongside time in a log-greatness 40-channel Mel-recurrence spectrogram, discarding undesirable qualities, at that point adding crosswise over recurrence. Further gradually differing counterbalances are expelled by a high pass channel at about 0.5 Hz.

At that point, a rough worldwide rhythm is evaluated via auto-connecting the beginning quality, applying an 'inclination window' which is a Gaussian on a log-time hub, then picking the period with the biggest windowed autocorrelation as the beat. Essentially, variety of the focal point of the inclination window somewhere in the range of 0.25s and 0.5s for example somewhere in the range of 240 and 120 beats for every moment (BPM) is utilized to get beat portions at various focuses in the metrical chain of command of the music. Subsequently utilized powerful programming to locate a beat times arrangement to upgrades both the beginning quality at each beat and the dividing between beats.
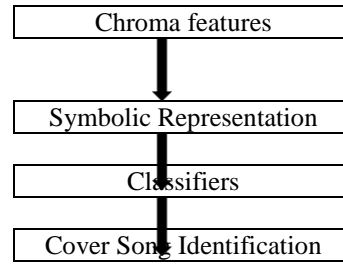

Chroma features
↓
Symbolic Representation
Classifiers
Cover Song Identification

Figure 1: The proposed method

B. Chroma Features

In the event that the beat following can distinguish a similar principle beat in various versions of a similar piece, at that point speaking to the sound alongside a period base characterized by the recognized beats standardizes away varieties in rhythm. Further record a solitary element vector for every beat, and utilize twelve component 'Chroma' highlights to catch together the prevailing note, regularly song just as the wide symphonious backup [8, 9]. Consequently the Chroma highlights record the force related with every one of the twelve semitones for instance piano keys inside one octave, however all octaves are collapsed organized. Calculating consonant highlights over beat-length sections shows up in [10].

Moderately than utilizing a rough mapping of FFT canisters to the Chroma periods they cover, utilize the stage subordinate inside each FFT receptacle both to recognize solid tonal segments in the range and to get a higher-goals gauge of the fundamental recurrence [11, 12]. Further this strategy help to expel non-tonal parts and improve recurrence goals past FFT receptacle level has comparative inspiration and effect to the sinusoid-demonstrating based preprocessing planned by [13]. By utilizing just parts up to 1 kHz Chroma highlights worked appropriately.

IV. SYMBOLIC REPRESENTATION

In this segment, we use chrome features for emblematic portrayal of tunes tests. As melody test have significant intra class varieties in every subgroup, utilizing regular information portrayal, saving these varieties is troublesome. Consequently, the proposed work is expect to utilize whimsical information handling called representative information examination which can protect the varieties among the information all the more adequately. In this work, emblematic portrayal has been adjusted to catch these varieties through component digestion by the utilization of an interim esteemed element vector as pursues.

Maximum and Minimum Representation

Let $[T_1, T_2, T_3, \ldots, T_n]$ be groups song classes say $C_j; j = 1, 2, 3, \ldots N$ here N represents number of classes and let $F_i = [d_{i1}, d_{i2}, d_{i3}, d_{i4}, \ldots, d_{im}]$ be the set of features. Let Min_{jk} and Max_{jk} be the minimum and maximum of the kth feature values obtained from class C_j.

$$Min_{jk} = \min(d_{ik}) \quad Max_{jk} = \max(d_{ik})$$

Where k=1 to n

Each class C_j is characterized by the use of interval valued features

$$([d_{j1}^-, d_{j1}^+], [d_{j2}^-, d_{j2}^+], \ldots, [d_{jk}^-, d_{jk}^+])_{\text{where}}$$

$d_{jk}^- = Min_{jk}$ and $d_{j1}^+ = Max_{jk}$ $d_{jk}^- = Min_{jk}$ and $d_{jk}^+ = Max_{jk}$

(1.1)

Now, a reference song demonstrating the entire class is designed by use interval type and is given by

$$RF_j = \left\{ [d_{j1}^-, d_{j1}^+], [d_{j2}^-, d_{j2}^+], \ldots, [d_{jm}^-, d_{jm}^+] \right\}$$

(1.2)

where c = 1, 2,… , of N.

Mean and Standard Deviation

Let $[T_1, T_2, T_3, \ldots, T_n]$ be song classes say $C_j; j = 1, 2, 3, \ldots N$ here N denotes number of classes and let $F_i = [d_{i1}, d_{i2}, d_{i3}, d_{i4}, \ldots, d_{im}]$ be the set of features the.

Let μ_{jk} and σ_{jk} be the mean and standard deviation of the feature values obtained from all the n samples of the class C_j.

$$\mu_{jk} = \frac{1}{n} \sum_{i=1}^{n} f_{ik}$$

$$\sigma_{jk} = \left[\frac{1}{n} \sum_{i}^{n} (f_{ik} - \mu_{jk})^2 \right]^{1/2}$$

Where k=1 to n

Each class C_j is characterized by the use of interval valued features

$$([d_{j1}^-, d_{j1}^+], [d_{j2}^-, d_{j2}^+], \ldots, [d_{jk}^-, d_{jk}^+])$$

(1.3)

Now, a reference song demonstrating the entire class is designed by use interval type and is given by

$$RF_j = \left\{ [d_{j1}^-, d_{j1}^+], [d_{j2}^-, d_{j2}^+], \ldots, [d_{jm}^-, d_{jm}^+] \right\}$$

(1.4)

where c = 1, 2,… , of N.

V. CLASSIFICATION

In this area we utilize the emblematic classifier for ordering the melodies. In arrangement model, a test of an obscure melody is depicted by a lot of m separations of sort fresh and contrasts it and the comparing interim sort highlights (separations) of the individual representative reference tests RFj put away in the knowledgebase to find out the proficiency.

Let $F_i = [d_{t1}, d_{t2}, d_{t3}, d_{t4}, \ldots, d_{tm}]$ be an dimensional vector describing a test song. Let $RF_c; c = 1,2,3,\ldots,N$ be symbolic feature vectors stored in knowledgebase. During classification process each kth distance (feature) value of the test sample is compared with the respective intervals of all the representativesto examine if the feature value of the test sample lies within them. The test sample is said to fit in to the class with which it has a maximum acceptance count A_c.

Acceptance count A_c is given by,

$$A_c = \sum_{k=1}^{m} C(d_{tk}, [d_{tk}^-, d_{jk}^+])$$

(1.5)

Where,

$$C(d_{tk}, [d_{tk}^-, d_{jk}^+]) = \begin{cases} 1 & if(d_{tk} \geq d_{jk}^- \text{ and } d_{tk} \leq d_{jk}^+) \\ 0 & \text{otherwise} \end{cases}$$

At the point when the database is huge, there is a likelihood for a test to have a similar most extreme acknowledgment check with at least two classes. Below such conditions we prescribe to determine the contention by the utilization of the accompanying comparability quantity which registers the similitude esteem among a test and every one of the clashing classes state jth class.

$$Total\_Sim(F_t, RF_j) = \sum_{k=1}^{m} C(d_{tk}, [d_{tk}^-, d_{jk}^+])$$

(1.6)

Here $[d_{jk}^-, d_{jk}^+]$ represents the kth feature interval of the jth conflicting class, and

A.

$$C(d_{tk}, [d_{jk}^-, d_{jk}^+]) = \begin{cases} 1 & if(d_{tk} \geq d_{jk}^- \text{ and } d_{tk} \leq d_{jk}^+) \\ \max\left(\frac{1}{1+|d_{tk} - d_{jk}^-|*\delta}, \frac{1}{1+|d_{tk} - d_{jk}^+|*\delta} \right) + 1 & \text{otherwise} \end{cases}$$

where δ is a normalizing factor (1.7)

VI. RESULT

We have experimented on created and real datasets in order to reveal the capability of proposed criteria. This method has been implemented in a Matlab R2013a using an Intel Pentium 4 processor, 2.99 GHz Windows PC with 1 GB of RAM. In this work, an attempt is made to compute Chrome feature with different datasets like covers30 and covers80. In addition, to reveal performances of Symbolic classifiers the experimentation is conducted on Min-Max Representation and Mean-Std Deviation Representation by varying the training samples. We select images randomly from the database and experimentation is conducted on database of 30 and 80 classes under changing preparation samples from 10 to 80 percentage for each class.

Figure 2 shows accuracy for Cover 30 songs under varying training samples for Min-Max Representation and Figure 3 shows accuracy for Cover 80 songs under varying training samples for Min-Max Representation. Figure 4 shows accuracy for Cover 30 songs under varying training samples for Mean-Std Representation and Figure 5 shows accuracy for Cover 80 songs under varying training samples for Mean-Std Representation. In all cases, when training percentage is 80 the system achievers maximum accuracy. The obtained results witnessed through maximum accuracy remarked in all cases. In view of subsequent appreciation and corroborated the supremacy of Classifier, the experimentation has been explored on database under varying size from 10 to 80 percent of database for all the cases. Figure 6 shows the comparison results of the proposed method. The proposed effort has shown that Min-Max Representation achieves maximum accuracy when compare to Mean-Std Deviation Representation.

Cover Song Identification through Symbolic Representation and Classifier
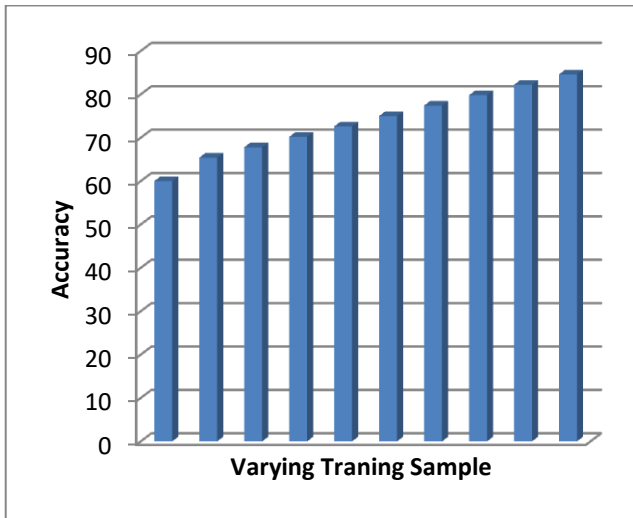

Figure 2: Shows the Accuracy of varying training samples for Cover 30 (Min-Max Representation)
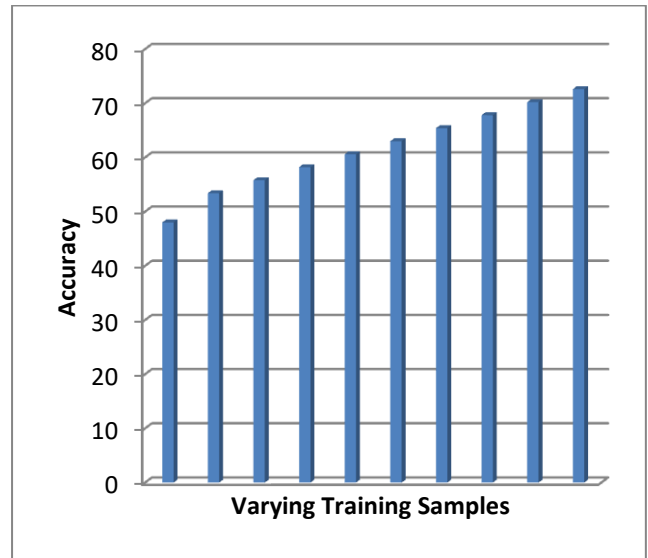

Figure 3: Shows the Accuracy of varying training samples for Cover 80 (Min-Max Representation)

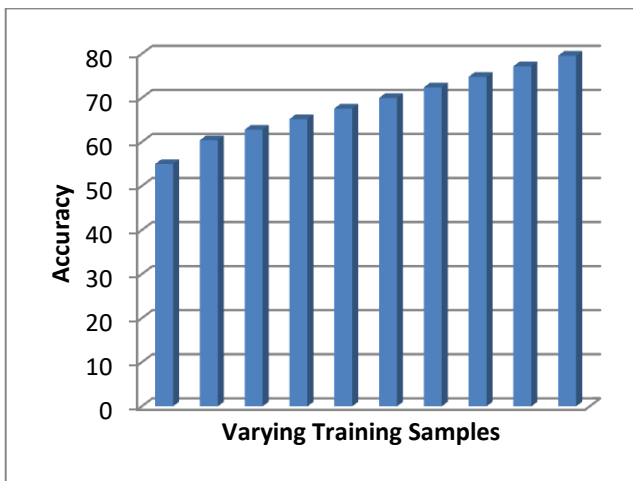**Figure 4: Shows the Accuracy of varying training samples for Cover 30 (Mean-Std Representation)**

Figure 5: Shows the Accuracy of varying training samples for Cover 80 (Mean-Std Representation)
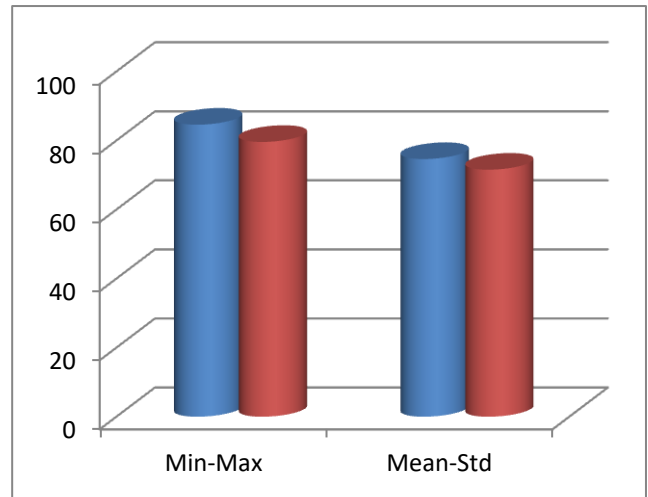

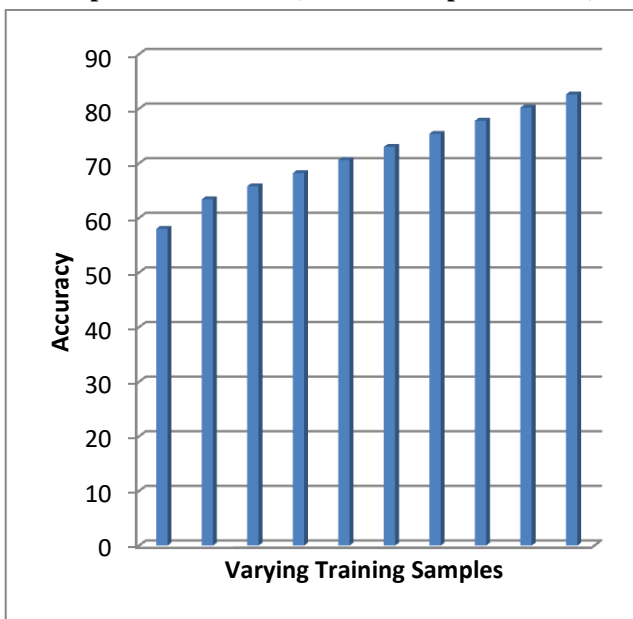**Figure 6: Shows the Accuracy by comparing with both the methods**

VII. CONCLUSION

In this paper, we made a fruitful endeavor to investigate the appropriateness of the ideas of emblematic information to Cover Song Detection. The recently proposed model has a capacity to catch the varieties of the highlights in preparing tests. The emblematic portrayal decreases the time taken to arrange a given test, as there is just a single delegate vector rather than n (preparing tests) number of agent vectors in the learning base. So as to research the adequacy and heartiness of the proposed strategy, we have directed broad trials on spread 30 and spread 80 dataset.

REFERENCES

1. J. Serr`a, E. G´omez, P. Herrera, and X. Serra. Chroma binary similarity and local alignment applied to cover song identification. IEEE Transactions on Audio, Speech and Language Processing, 16:1138–1151, 08/2008 2008.
2. J. V. Balen, D. Bountouridis, F. Wiering, and R. Veltkamp. Cognition-Inspired Descriptors for Scalable Cover Song Retrieval. In ISMIR'14, pages 379–384, 2014.

3. B. Martin, D. G. Brown, P. Hanna, and P. Ferraro. BLAST for Audio Sequences Alignment: A Fast Scalable Cover Identification Tool. In ISMIR'12, number Ismir, pages 529–534, 2012.

4. M. Khadkevich and M. Omologo. Large-Scale Cover Song Identification Using Chord Profiles. In ISMIR'13, pages 233–238, 2013.

5. Y. Lu and J. E. Cabrera. Large scale similar song retrieval using beat-aligned chroma patch codebook with location verification. In SIGMAP'12, pages 208–214, 2012.

6. E. J. Humphrey, O. Nieto, and J. P. Bello. Data Driven and Discriminative Projections for Large-Scale Cover Song Identification. In ISMIR'13, pages 149–154, 2013.

7. J. Salamon, J. Serr`a, and E. G´omez. Tonal representations for music retrieval: From version identification to query-by-humming. International Journal of Multimedia Information Retrieval, special issue on Hybrid Music Information Retrieval, 2:45–58, 2013.

8. A. Degani, M. Dalai, R. Leonardi, and P. Migliorati. A Heuristic for Distance Fusion in Cover Song Identification. In WIAMIS'13, pages 1–4, 2013.

9. T. Fujishima. Realtime chord recognition of musical sound: A system using common lisp music. In Proc. ICMC, pages 464–467, Beijing, 1999.

10. M. A. Bartsch and G. H. Wakefield. To catch a chorus: Using chroma-based representations for audio thumbnailing. In Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, Mohonk, New York, October 2001.

11. N. C. Maddage, C. Xu, M. S. Kankanhalli, and X. Shao. Content-based music structure analysis with applications to music semantics understanding. In Proc. ACM MultiMedia, pages 112–119, New York NY, 2004.

12. F. J. Charpentier. Pitch detection using the short-term phase spectrum. In Proc. ICASSP-86, pages 113–116, Tokyo, 1986.

13. T. Abe and M. Honda. Sinusoidal model based on instantaneous frequency attractors. IEEE Tr. Audio, Speech and Lang. Proc., 14(4):1292–1300, 2006.

14. Ellis, Daniel P.W., Cotton, Courtenay V. "The 2007 LabRosa Cover Song Detection System," Music Information Retrieval Evaluation Exchange, 2007.

15. M.-L. Zhang and Z.-H. Zhou. ML-KNN: A Lazy Learning Approach to Multi-Label Learning. Pattern Recogn., 40(7):2038-2048, 2007.

16. X. Hao, X. Tao, C. Zhang, and Y. Hu. An Effective Method to Improve kNN Text Classifier. In Eighth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, volume 1, pages 379-384, 2007.

17. K. Z. Mao, K. C. Tan, and W. Ser. Probabilistic Neural- Network Structure Determination for Pattern Classification. IEEE Transactions on Neural Networks, 11:1009-1016, 2000.

18. P. K. Patra, M. Nayak, S. K. Nayak, and N. K. Gobbak. Probabilistic Neural Network for Pattern Classification. IEEE Proceedings of the 2002 International Joint Conference on Neural Networks, vol. 2:1200 - 1205, 2002.

19. V.N.Vapnik, "The Nature of Statistical Learning Theory", Springer-Verlag, New York, 1995.

20. C. J. C. Burges, "A tutorial on support vector machines for pattern recognition", Data Mining and Knowledge Discovery, 2 (2), pp. 121–167, 1998.

21. Schatzman, James C, "Accuracy of the discrete Fourier transform and the fast Fourier transform". SIAM Journal on Scientific Computing. 17: 1150–1166, 1996.

22. Mei Jiansheng, Li Sukang. Tan Xiaomei," A Digital Watermarking Algorithm Based On DCT and DWT", Proceedings of the 2009 International Symposium on Web Information Systems and Applications (WISA'09) Nanchang, P. R. China, May 22-24, 2009, pp. 104-107.

23. An Overview of Non-contact Photoplethysmography, Peck Y S Cheang and Peter R Smith, ELECTRONIC SYSTEMS AND CONTROL DIVISION RESEARCH, 2003.

24. Frigo, M. and S. G. Johnson, "FFTW: An Adaptive Software Architecture for the FFT,"Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, Vol. 3, 1998, pp. 1381-1384.

25. Zanaty EA, Afifi A, Khateeb RE. Improving the accuracy of support vector machines via a new kernel functions. Int J IntellComputSci 2009;1:55–67.

26. Zanaty EA, Aljahdali S, Cripps RJ. Accurate support vector machines for data classification. Int J Rapid Manuf 2009;1(2).

27. EktefaMohammadreza, Sidi Fatimah, Ibrahim Hamidah, JabarMarzanah, Memar Sara. A comparative study in classification techniques for unsupervised record linkage model. J ComputSci 2011;7:341–7.

28. SonkambleBalwant A, Doye D. A novel linear-polynomial kernel to construct support vector machines for speech recognition. J ComputSci 2011;7:991–6.

29. Zanaty EA, Afifi A. Support vector machine (SVMs) with universal kernel. ApplArtif. Intell. 2011;25(7):575–89.

30. DeCoste D, Mazzoni D. Fast query optimized kernel machine classification via incremental approximate nearest support vec-tors. In: Proceedings of the 20th international conference on machine learning, Washington, ICML 2003; 2003. p. 115–22.

31. Scho¨lkopf B, Smola AJ. Learning with kernels. Cambridge, MA: The MIT Press; 2002.

32. Scho¨lkopf BS, Mika CJ, Burges P, Knirsch RR, Muller KG. Input space versus feature space in kernel-based methods. IEEE Trans Neural Network 1999;10(5):1000–17.

33. Burges CJ, Scho¨lkopf B. Improving the accuracy and speed of support vector machines. Neural Inform Proc Syst 1997;9:375–81.

34. Burges CJ. A tutorial on support vector machines for pattern recognition. Data Min Knowledge Discovery 1998;2(2):121–67.

35. Scho¨lkopf B, Burges C, Smola A. Advances in kernel methods support vector learning. MIT Press; 1998.

36. Taylor SJ, Bartlett PL, Willianmson RC, Anthony M. Structural risk minimization over data-dependent hierarchies. IEEE Trans Inf Theory 1998;5:1926–40.

37. Making JT. Large-scale SVM learning practical. In: Advances in kernel methods-support vector learning. Cambridge: MIT Press; 1999. p. 42–56.

38. O. Chapelle, V. Vapnik, O. Bousquet, S. Mukherjee, Choosing multiple parameters for support vector machines, Machine Learning 46 (1–3) (2002) 131–159.

39. Suresh Yerva, Smita Nair and Krishnan Kutty,‖ Lossless Image Compression based on Data Folding,‖IEEE, pp. 999-1004, 2011.

40. T. Tsai, T. Prätzlich, M. Muller, "Known-artist live song id: a hashprint approach", International Society for Music Information Retrieval Conference, 2016.

41. PremSeetharaman, FatemehPishdadian, Bryan Pardo, "Music/Voice separation using the 2D fourier transform", Applications of Signal Processing to Audio and Acoustics (WASPAA) 2017 IEEE Workshop on, pp. 36-40, 2017.

AUTHORS PROFILE

D. Khasim Vali, received the B.Tech in Computer Science and Engineering from the Jawaharlal Nehru technological university, Hyderabad, Andhra.Pradesh, India in 2005 then received the M.Tech in Master of Engineering Management from Visvesvaraya Technological University, Belgaum, Karnataka, India in 2008 and enrolled for PhD in Computer Science under Visvesvaraya Technological University, Belgaum, Karnataka, India in the year 2013. Since August 2007, he has been with department of Computer Science and Engineering, Vidyavardhaka College of Engineering, Mysur, Karnataka, India. He is currently working as an Associate Professor at the Department of Computer Science and Engineering. , Vidyavardhaka College of Engineering, Mysuru, Karnataka, India. He has 21 international publications to his credit. His research interests include artificial intelligence, Machine Learning, and Signal processing. He is a member of ISTE and IETE.

Nagappa U. Bhajantri, did his BE from UBDT, Davangere, Karnataka, India, in 1991. M.Tech from Indian Institute of Technology, Delhi, India in 1999 and PhD from the University of Mysore, Karnataka, India in 2007. He has served at Sri Jayachama Rajendra College of Engineering, Mysore, Karnataka, India from 1994–2010. His research interests include image, video processing and also melody recognition. He has 16 international publications to his credit. He has produced three PhDs successfully and four more candidates currently working. He is currently working as a Professor at Government Engineering College, Chamarajanagara, Karnataka, India.