Journal of Information Systems Engineering and Management
2025, 10(31s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Swaram Extraction from Veena Tunes using Deep Learning Algorithms

Dr. S. Mythili[1], Dr. M. Rajesh Babu[2], Dr. G. Naveen Sundar[3], Dr. S. Uma[4], Dr. P. Thangavel[5], S. Anitha[6]
[1]*Associate Professor, Department of CSE, United Institute of Technology, Coimbatore, India*
Email: mythili@uit.ac.in
[2]*Professor & Associate Dean, School of Computing, Rathinam Technical Campus, Coimbatore, India*
Email: drmrajeshbabu@gmail.com
[3]*Associate Professor, Division of CSE, Karunya Institute of Technology and Sciences, Coimbatore, India*
Email: naveensundar@karunya.edu
[4]*Professor, PG In charge, Department of CSE, Hindusthan College of Engineering and Technology, Coimbatore, India*
Email: druma.cse@hicet.ac.in
[5]*Assistant Professor [SG], Department of IT, Government College of Engineering, Erode, India*
Email: thangsirtt@gmail.com
[6]*PG Student, Department of CSE, Hindusthan College of Engineering and Technology, Coimbatore, India*
Email: anithasaravanacsc@gmail.com
Corresponding author email id: mythili@uit.ac.in

ARTICLE INFO

Received: 26 Dec 2024

Revised: 14 Feb 2025

Accepted: 22 Feb 2025

ABSTRACT

Music is one of the fast-growing industries in today's world and faces many difficulties. Indian Classical music, too, is different from other western music patterns and difficult to learn or identify. A lot of work has been done before using traditional approaches and expertise to identify Swaram(notes) with the help of standard features such as arohana, avarohana, pakad, gamak, vaadi, savandi etc. but with the help of new features such as chromagrams. Audio information retrieval (AIR) is a field with potential applications in automatic annotation, music recommendation, as well as music tutoring and accuracy verification systems. Extracting the Swaram (notes), or melodic style, of improvisational Classical music is a challenging problem in AIR due to the music's melodic variation and inconsistent temporal spacing. In this project, we propose a deep learning-based approach to Swaram (notes) recognition. Deep learning system is proposed for extracting information from audio data with temporal variation. Our method makes effective use of long short-term memory based recurrent neural networks to efficiently pre- possess and learn temporal sequences in music data (LSTM-RNN). Swaram identification is a sequence classification task in which each Swaram is treated as a class and the notes produced by prevailing melody estimation are treated as words. We train and test the network on smaller sequences sampled from the original audio while the final inference is performed on the audio as a whole.

Keywords: *LSTM-RNN, AIR, ANN, Deep Learning*

INTRODUCTION

Indian classical music is known for its perfect technical soundness and well-defined structure known as swarms. Each swaram is based on some specific combination of swara (notes). Any swaram should have at least five notes out of seven and it is also possible that two swaram has same notes but the aarohan, avrohan and pakad is different so one can identify it properly. Expert person can understand the swaram very easily, but for learner it is very difficult to classify and identify the swaram. So, this method is helpful forprofessional and non- professional one. Ensemble of deep learning models are proposed to achieve better prediction performance when compared to that of the individual models [1]. Proposed work is related to classification and analysis of Instrumental Indianclassical music. We are using MATLAB tool for processing music segment and find out information related to swaram analysis and classification. Music Information Retrieval toolbox (MIR) [3] is also helpful to find out the features for comparison. This software is widely used in western music and now days implemented in Indian Classical music. Work focused on four swarams namely Bhairav, Bhairavi, Todi andYaman and the selected instrumental music is mixed polyphonic

to find out the spectral and temporal features likebrightness, RMS energy, spectral flux spectrum chromogram, histogrametc. For classification we preferred KNN and SVMclassifier [3], [4] Swaras are the frequency generated by instrumentally or vocally. Actually these seven swaras represents the absolute frequencies ratio with respect to each other and these are very similar to SOLFEGE in western music The seven swaras are namely: Shadja (Sa), Rishabh (Re), Gandhara (Ga), Madhyama (Ma), Panchama (Pa), Dhaivata(Dh), and Nishad (Ni).Out of seven, two swaras i.e Sa and Pa has only pure form while other five has both pure and impure form in structural elements of swaram. The purpose behind this work is to design a computer-based education of Indian classical music for everyone.

Following are the basic terms related to Indian classical music.

Swaram:

Indian music is famous for its swarams specialty. Swaram represents the color of emotion and it has some specific melodic phrases. Swaram has definite pattern of notes (swaras). Two swarams may be same notes but the patternof distribution is different in case of each swaram that is why two swarams have same notes but they tuned differently.

Aarohan ,Avrohan and Pakad

Aarohan and Avrohan is the ascending and descending progression of swaras respectively. A pakad, which is a brief group of swaras in a swaram that serves as a mark for the swaram, is often called out in any classical performance.

Vadi , Samvadi and Jati

The legacy of Indian Carnatic Classical music has its roots back to 1500BC. Swarams form the backbone of Indian Classical music. Before the technological era, musicians usedtheir hearing to recognize notes and swarams in a composition. Musicians then were able to understand the tonal differences even lesser than 0.2Hz. Carnatic music is self-possessed with seven diverse keys called Swaras viz.

Sa (Shadja) Ri (Rishabha)

Ga (Gandhara) Ma (Madhyama)Pa (Panchama) Dha (Dhaivata) Ni (Nishada)

The different combinations of Swaras with adherence to specific rules make up the seventy-two Melakartha swarams. Each swaram is composed of defined notes and depicts a specific mood. There are attributes related to each swaramthat enhances its feel and emotions. Some of these attributes are Arohana and Avarohana (ascending and descending progressions), Gamaka (ornamentation of different notes ina pattern), Vadi (root swara of the swaram and the most important note in a swaram), Samavadi (the second mostimportant note in a swaram), Tala (rhythmic pattern) and Samay (specific time in which a swaram shows its dominance and makes the recognition of that swaram easy). The seventy-two Melakartha swarams constitute the Indian Carnatic music. These swarams are the parents of all other sub swarams in Carnatic music. Each swaram is classified based on its key combination, which is unique. Swarams that are derived from Melakartha swarams are called Janya swarams. There are numerous Janya swarams foreach Melakartha swaram. The number of notes present and their aro hana and avarohana patterns are the main differences between Melakartha swarams and Janya swarams. Janya swarams require should consist of minimum of five notesfrom a Melakartha swaram. Moreover, unlike their parent Melakartha swarams, arohana and avarohana patterns need not be the same for Janya swarams.

Carnatic swaram recognition helps in identifying the swaramsin a song, which would highly assist Musicologists and aspiring musicians. Swaram recognition also helps in filteringsongs according to their swarams. Computational musicologyis an emerging and trending research area. Many works havebeen carried out in swaram identification in Indian Classical music. Different attributes that are used for swaram recognition are usage of notes, arohana and avarohana patterns, gamaka, pakad, vadi and time. In1, the frequencies in the song were captured at specific intervals using Pitch Class Distribution methods [PCD]. The voice of the singer was isolated from the orchestration using separation algorithm and segmented using segmentation algorithm. Then the singer was identified to get their fundamental frequency, after which, string matching was usedto identify the notes and thus the swaram.

In this project, the frequency was considered relative to the fundamental frequency of the singer. Pitch distribution method was the used method for swaram identification. However, the problem with this method was that in most of the attempts, voice of the singer had to be isolated from the song as the voice of instruments created bafflements in the process. Though this process yielded reasonably good results,the process was error prone.

Vadi, Time and Pakad were used as parameters for swaram identification in Hindustani music in2. The proposed method used fuzzy logic to find the interrelationship between the swaras. In this paper, importance is given to Samay (Time) when the swaram is sung. The author claimed that the mood of each swaram shows its prominence at a particular Samay which was used as the attribute for classification, converted models into a feature matrix to represent it in vector space. A model was created in this manner was then used for classification of swarams. Many algorithms were used to create the model and test it, among which, the Naive Bayesianclassifier was found to be giving the best results when compared with other algorithms.

A system to identify swaras in each Carnatic song using pitch distribution was proposed in [6]. Talam was given the most importance and was identified by the intensity of the sound. The starting point of the segment was filtered by the sound segment with the highest intensity and the end with thelower intensity. These segments were then mapped to their corresponding swaras.

Swaram identification was carried out in Indian Classical Music using swara intonation as a part of the study conductedin6. A database with the frequencies of specific swarams sungby four artists was created. Polyphonic melody extractor was used to find the fundamental frequency. The Peak (most likely position), mean position, standard deviation and overall probability were selected as features. Test result accuracy of 80% was achieved in this manner.

Arohana and avarohana patterns were used for swaram identification in [8], wherein, the features used as parameters forswaram identification were swara combination, number of swaras used in the swaram, vakra pairs in Arohana and Avarohana and swara combinations. Neural networks were used asthe algorithm for classification.

SYSTEM ANALYSIS:

EXISTING SYSTEM:

The Swaram detection is performed on the musical file from which features are extracted. The following classifiers are used previously for Swaram detection:

- **C4.5 classifie**r

C4.5 classifier builds decision trees from a set of training datausing the concept of information entropy. Each training sample si consists of a p-dimensional vector x1,i , x2,i,.....,xp,I where xj represent attributes or features of the sample, as well as the class in which si falls. At each node

of the tree, C4.5 chooses the attribute of the data that most adequately splits its set of samples into subsets enriched in one class or the other.

- **Bayesian classifier**

A Bayesian network is a probabilistic graphical model that represents a set of random variables and their conditional dependencies via a directed acyclic graph (DAG). Each node is associated with a probability function that takes as input a particular set of values for the node's parent variables and gives the probability of the variable represented by the node.

- **Random Forest classifier**

Random forests (RF) are a combination of tree predictors and uses random selection of features to split each node growing an ensemble of trees and letting them vote for the most popular class. To grow these ensembles, often random vectors are generated that govern the growth of each tree in the ensemble. After a large number of trees is generated, they vote for the most popular class.

- **K-star classifier**

K-star or K* is an instance-based classifier. The class of a testinstance is based on the training instances similar to it, as determined by some similarity function. The K-star algorithmuses entropic measure, based on probability of transforming an instance into another by randomly choosing between all possible transformations.

- Bayesian net
- Naive Bayes
- Support vector machine (SVM)

- K-NN
- Decision table
- Random forest
- Multi-layer perceptron
- PART

DISADVANTAGES

- Incorrect extraction of pitch.
- Manual detection of tonic
- Only the static work is done through identification
- Lack of a database and the incorrect extraction offeatures results in an incorrect result.
- The SVM classifier is difficult to handle scale andmultiple instruments.
- K-NN may cause problems with gammakas and pitchextraction.

PROPOSED SYSTEM

Identifying the Swaram form a given polyphonic music signal was attempted in this paper. A segment from a given song,with duration of 30 seconds, was considered as input.

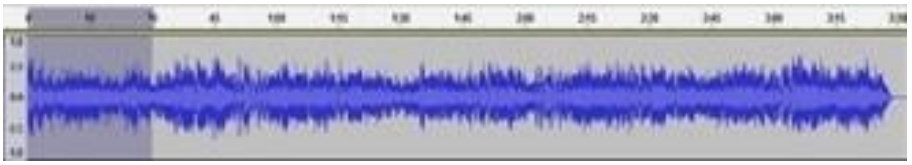


Figure1. Audio signal with 30 seconds selected segment

After preprocessing, the signal was considered for comparison and classification for Swaram identification.

In this study, we provide an approach to Swaram recognition based on deep learning. Our method makes use of effective prepossessing and recurrent neural networks based on long short-term memory to learn temporal structures in music data (LSTM-RNN).

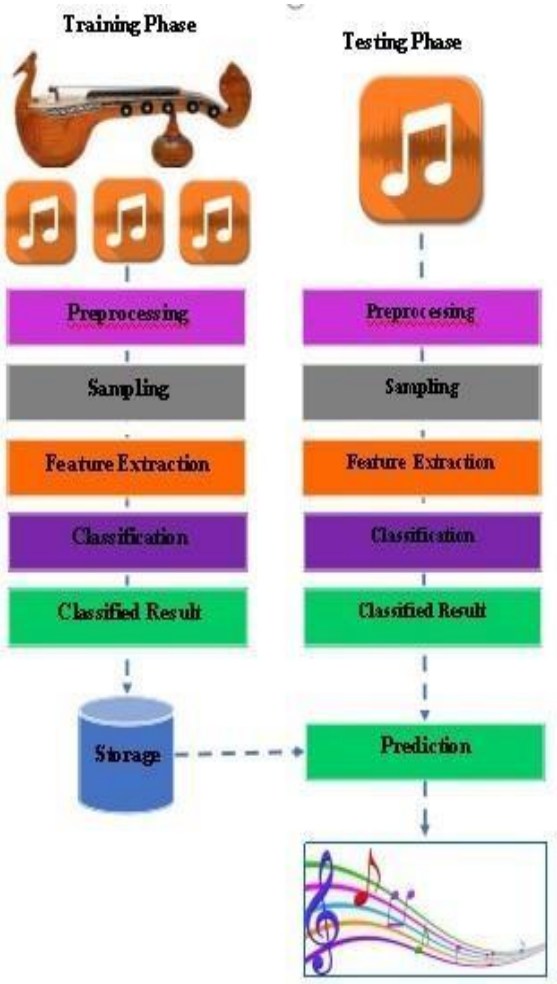
Swaram identification is reformulated as a sequence classification task where each Swaram is treated as a class and the notes assimilated by prevalent melody estimation are treated as words.

ADVANTAGES

Our approach in tackling the problem of Swaram recognition, retrieving melodic patterns from a given music data base that are closely related to the presented query sequence, it achieves an accuracy of 88.1% and 97%. Also, it enables the relatively fast prediction of Swaram.

- It can handle scale and multiple instruments.
- Low Computational overhead.
- Gentle extraction of pitch.

ARCHITECTURAL DIAGRAM:



SYSTEM REQUIREMENTS

Hardware Requirements

64-Bit MATLAB, Simulink and Polyspace Product Families

Operating Systems: Windows 10

Processors : Any Intel or AMD x86-64 processor **Disk Space** : Minimum 2.6 GB of HDD space for MATLAB only, 4-6 GB for a typical installation, 500GB **RAM** : 4 GB

Software Requirements

Matlab 2018b

MATLAB is a high-performance language for technical computing. It integrates computation, visualization, and programming environment. Furthermore, MATLAB is a modern programming language environment: it has sophisticated data structures, contains built-in editing and debugging tools, and supports object-oriented programming. These factors make MATLAB an excellent tool for teaching and research.

SIMULINK

Simulink is a simulation and model-based design environment for dynamic and embedded systems, integrated with MATLAB. Simulink, also developed by MathWorks, is a data flow graphical programming language tool for modelling, simulating and analyzing multi-domain dynamic systems. It is basically a graphical block diagramming tool with customizable set of block libraries.

- It allows you to incorporate MATLAB algorithms into models as well as export the simulation results into

SYSTEM DESIGN

Modules Description

1. Veena tune Annotation

Training: Collection of tunes

Testing: Single Tune

2. Pre-processing

Pre-processing is the second step of methodology; in thisstep we take all the tunes and first converts these files into .WAV files. After conversion we cut each file into a specific duration i.e., approximates 10 sec so that we can get information about each and every note properly. After pre –processing we fed the data into feature extraction.

- Data is cleaned by removing speech, applause and narration from every file.
- All the audio files are converted to .wav format with16-bit PCM encoding. Converted the entire files mono format using only single channel. All the 10sec files are encoded at the same sampling rate of 22050. Music file length was decided to have 10 sec and sofile clipping and appending was done to maintain thesame length.

3. Sampling Frequency

In this module all the music files are sampled at a definedrate and segmented into frames.

An example of frame and sample calculation is given below.

- Length of music file = 10s
- Sampling rate = 22050
- Total samples = 10*22050 = 220500.
- Hop length (number of samples in frame) = 512 (aselectable parameter in librosa)
- Frame length = sample rate * hop length = (1/22050) *512

= 23.2ms which fall under standardframe length of 20–40ms.

- Total number of frames in music file = 2205000/512 = 430.Spectral Centroid gives the mean frequency of each frame inthe music file.

4. Feature Extraction

Extraction of features means that the specialty or pattern thatis repeated in that signal will be observed and extracted as a feature. It will help to distinguish between different tune, as each Swaram has its own characteristics and the pattern through which it is identified. Different tunes are composed of different Swaras and their variations. A frequency is associated with each variation of Swaram. So, plotting of spectral centroid tell us how the frequency is varying. SpectralBandwidth gives the frequency bandwidth available in every frame of music file. Chromagram gives information about thepitch classes in each frame of music file. Energy and RMSE gives us the energy of the signal for each frame. This gives the information about loudness of the signal. MFCC are smallset of features which concisely describe the overall shape of the spectral envelope. Mel scale is derived based on the human perception of audio frequencies. CQT is similar to Fourier transform but similar to Mel-scale uses logarithmically spaced frequency axis.

Features considered are given below.

1. Spectral centroid
2. Spectral bandwidth
3. Mel-spectrogram
4. Mel frequency cepstral coefficients.
5. Chroma

5. Classification

Swaram recognition as a sequence classification task performed using an LSTMRNN based architecture with attention. During the Classification process, we sample from a qualified probability distribution to generate new music pieces. It has been validated that recurrent neural networks (RNN), particularly long-term memory networks (LSTMs), can accurately forecast time series data.

A type of artificial neural network called a recurrent neural network (RNN) is one in which correlations between nodes create a graph along a series. This marks it likely to see time-dependent behaviour for a time series. They commenced by means of three GRU layers layered on top of one another, each with 512 hidden neurons, but later decided to experiment with LSTM layers and change the number of hidden neurons and the number of layers in order to better understand what is best accomplished.

6. Prediction

The joint Swaram prediction system achieves a new state-of-the-art 98.9% accuracy for Swaram prediction on 30s length tunes.

7. Performance Evaluation

We use both frame and note based metrics to assess the performance of the proposed system. Frame-based evaluations are made by comparing the transcribed binary output and the MIDI ground truth frame-by-frame. For note- based evaluation, the system returns a list of notes, along withthe corresponding pitches, onset and offset time. We use the Fmeasure, precision, recall and accuracy for both frame and note based evaluation. Formally, the frame-based metricsare defined as:

$$\mathcal{P} = \sum_{t=1}^{T} \frac{TP[t]}{TP[t] + FP[t]}$$

$$\mathcal{R} = \sum_{t=1}^{T} \frac{TP[t]}{TP[t] + FN[t]}$$

$$\mathcal{A} = \sum_{t=1}^{T} \frac{TP[t]}{TP[t] + FP[t] + FN[t]}$$

$$\mathcal{F} = \frac{2 * \mathcal{P} * \mathcal{R}}{\mathcal{P} + \mathcal{R}}$$

where TP[t] is the number of true positives for the event at t, FP is the number of false positives and FN is the number of false negatives. The summation over T is carried out over theentire test data. Similarly, analogous note-based metrics can be defined. A note event is assumed to be correct if its predicted pitch onset is within a ±50 ms range of the ground truth onset.

ALGORITHM DESCRIPTION:

LONG SHORT-TERM MEMORY (LSTM)

An artificial recurrent neural network (RNN) architecture [1] called long short-term memory (LSTM) is utilised in deep learning. LSTM features feedback connections as opposed to typical feedforward neural networks. It can analyse whole data sequences in addition to single data points (like photos) (such as speech or video). For instance, LSTM can be used for applications like linked, unsegmented handwriting identification [2], speech recognition [3][4], anomaly detection in network traffic, and intrusion detection systems (intrusion detection systems).

A cell, an input gate, an output gate, and a forget gate make up a typical LSTM unit. The three gates control the flow of information into and out of the cell, and the cell remembers values across arbitrary time intervals.

Since there may be lags of uncertain length between significant occurrences in a time series, LSTM networks are well-suited to categorizing, processing, and making predictions based on time series data. To solve the vanishing gradient issue that can arise when training conventional RNNs, LSTMs were created. The advantage of LSTM over RNNs, hidden Markov models, and other sequence learning techniques in many applications is their relative insensitivity to

gap length.

OPERATION OF THE SYSTEM

Sample data for training the system:

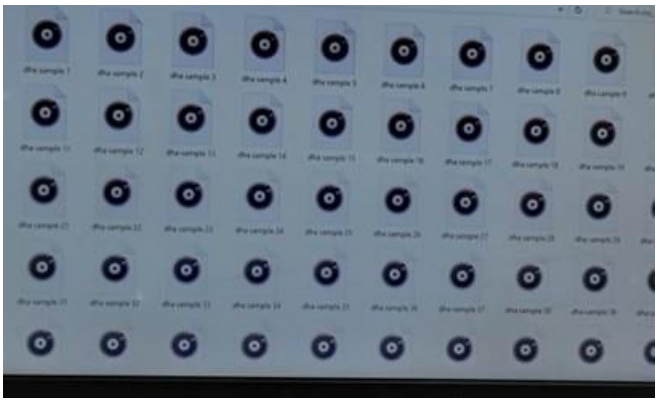


Figure 2. Sample data

These are the sample data collected for training the system.

User interface:

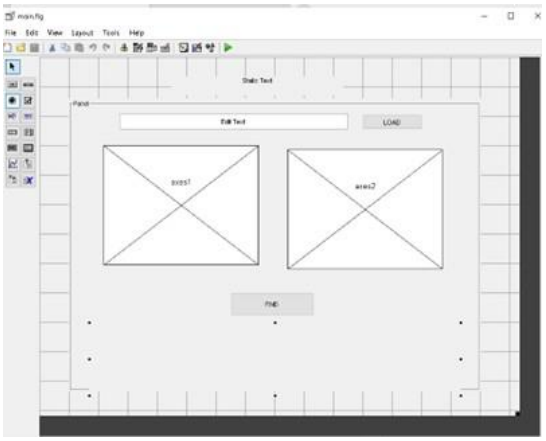
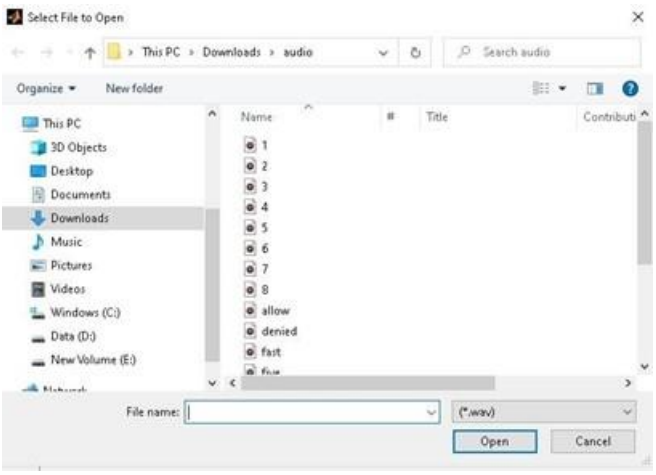


Figure 3. User interface

This is the screen that will be displayed while staring theapplication. Click on the “LOAD” button to load the input.

These are the sample audio file for input.Output Screen:



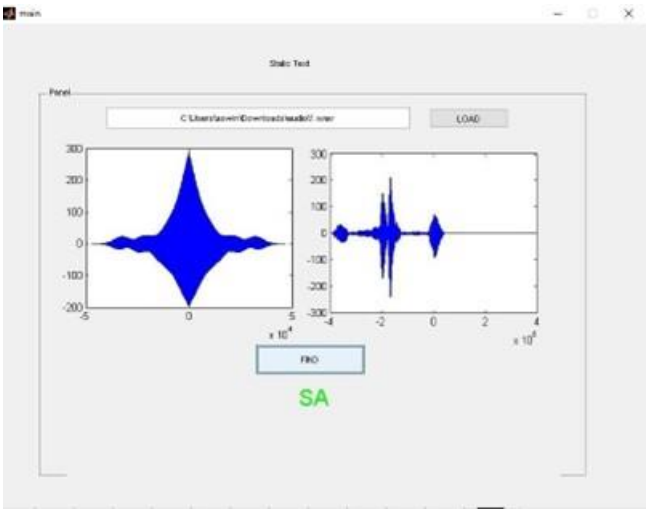


Figure 4. Output Screen

After selecting the input click on the "FIND" button in the bottom to display the output. The output will be display as shown in the Figure 4.

CONCLUSION AND FUTURE WORKS

In this paper, we have presented an innovative approach to the Swaram recognition issue. We can show the effectiveness of our strategy in addressing the Swaram recognition challenge through a variety of tests and validation. We also add sequence ranking as a brand-new Swaram identification subtask. As complementary information, we provide a large number of query-retrieved sub sequence combinations that display how effective this method is for probing databases for related sequences. The suggested hierarchical system can be extended in the future to handle more tasks requiring lengthy temporal sequences, and the data augmentation and

REFERENCES

[1] Nandhini, T., Rajesh Babu, M., Rajalakshmi, S., Rajasekar, M., and Sivakumar, T., "Rectal Fundus Image Recognition Using Deep Learning Ensemble CNN Models", Journal of Information Systems Engineering and Management, Vol.10, No. 26s, 2025.

[2] Dipti Joshi, Jyothi Pareek and Pushkar Ambatkar, "Indian Classical Raga Identification using Machine Learning," International Semantic Intelligence, 2021.

[3] 3. R.Shridhar and T. V. Geetha, "Swaram identification of Carnatic music information retrieval," International Journal of Recent Trends in Engineering, vol. 1, no. 1, pp. 571-574, 2019.

[4] A. Bhattacharjee and N. Sriniwasan, "Hindustani Swaram representation and identification: A transition probabilitybased approach," International Journal of Mind, Brain and Cognition, vol. 2, pp. 65-93, 2011.

[5] S.Shetty and K. K. Achary, "Swaram mining of Indian music by extracting Arohana-Avarohana pattern," International Journal of Recent Trends in Engineering, vol. 1, no. 1, pp. 362-366, May 2009.

[6] G.Pandey, C. Mishra, and P. Ipe, "Tansen: A system for automatic Swaram identification," in Proc. 1st Indian International Conference on Artificial Intelligence, Hyderabad, India, 2003, pp. 1350- 1363.

[7] Kyogu Lee, “Automatic chord recognition from audiousing enhanced pitch profile:, ICmC 2006.
[8] Parag Chordia, "Understanding Emotion in Raag: An Empirical Survey of Listener Responses." In Proc. of the 2007, International Computer Music Conference (ICMC).
[9] Hirose Y, Yamashita K Y and Hijiya S (1991), Back- propagation algorithm which varies the number of hidden units, Neural Networks, 4, pp 61-66.
[10] Alai de Cheveigne, “YIN, a Fundamental frequency estimator for speech and music”, Journal of Acoustical Society of America., Vol. 111, No. 4, April 2002.
[11] Krishnaswamy, A, “Application of pitch tracking to South Indian classical music”, Applications of Signal Processing to Audio and Acoustics, 2003 IEEE Workshop on 19-22 Oct. 2003.
[12] S. Shetty and K. K. Achary, “Swaram mining of Indian music by extracting Arohana-Avarohana pattern,” International Journal of Recent Trends in Engineering, vol. 1, no. 1, pp. 362-366, May 2009.