Automated Raaga Recognition in Carnatic Music Using Light Weight CNN-BiLSTM Attention Model

By Group 8

A Venkata Satya -CB.EN.U4AIE22005
Lakshmi KN -CB.EN.U4AIE22023
K Hemavardhan Reddy -CB.EN.U4AIE22026
K V Vamshidhar Reddy -CB.EN.U4AIE22028

Table of Contents

INTRODUCTION AND BACKGROUND

- Music plays a significant role in people life.
- In India there are 2 forms of classical music. Hindustani music and Carnatic Music.
- Both revolve around the concept of Raaga and Carnatic music is based on Seven swaras
- Raaga is a set of rules defined for making compositions..
- Raagas are defined by the distribution of these swaras, arohanam (asecending movement of the Swaras), Avarohanam (Descending movement of the swaras), Gamakas.
- Raagas are classified into Melakartha(parent Raaga)and Janya Raaga (child Raaga)

LITERATURE REVIEW

TITLE	AUTHOR	MODEL DESCRIPTION	DATASET	PERFORMANCE	SOURCE
Classification of Indian Classical Carnatic Music Based on Raga Using Deep learning	Siji John Et al	CNN model is implemented using Keras with TensorFlow as the backend, following a Sequential architecture with convolutional, pooling, flatten, and fully connected layers.	Comp Music Dataset	94 percent for the five typical ragas under consideration.	https://ieeexplore.ieee.org/document/9332482
ASurvey of Raaga Recognition Techniques and Improvements to the State-of-the-Art	Gopala Krishna Koduri Et al	A kNN classification framework is adopted where several values of k are tried.	Own Dataset consisting of 170 tunes across 10 raagas, including vocal and instrumental performances from 31 artists.	76.5%,	https://www.researchgate.net/publication/264885437_A_Survey_of_Raaga_Recognition_Techniques_and_Improvements_to_the_State-of-the-Art
Raga Recognition in Indian Classical Music using Deep Learning	Devansh P Shah Et al.	The proposed model for raaga recognition is a hybrid CNN-LSTM architecture, where CNN extracts features from spectrogram images and LSTM captures temporal dependencies across sequences of these features.	Comp Music Dataset	98..98%	https://www.researchgate.net/publication/350553712_Raga_Recognition_in_Indian_Classical_Music_Using_Deep_Learning

TITLE	AUTHOR	MODEL DESCRIPTION	DATASET	PERFORMANCE	SOURCE
Multimodal Deep Learning Architecture for Hindustani Raga Classification	Stella Et al	The proposed classifier is a deep neural network based on multimodal learning, combining audio signal processing with metadata analysis for raga identification.	CompMusic Art Indian music dataset	98.22%.	https://www.proquest.com/docview/2841538396?pq-origsite=gscholar&fromopenview=true&sourcetype=Scholarly%20Journals
Machine Learning Based Indian Raga Classification and Detection	Bhagyalakshmi R Et al	HMM model which undergone learning using the Baum Welch procedure.	GTraagDB	90%	https://ieeexplore.ieee.org/abstract/document/10673605
Explainable Deep Learning Analysis for Raga Identification in Indian Art Music	Parampreet Singh Et al	Trained a CNN-LSTM model to perform multi-class classification for N = 12 Raga classes	Own Dataset	f1-measure of 0.89	https://arxiv.org/abs/2406.02443

OBJECTIVE

- Develop an efficient raaga recognition model that :
- Is a light weight model.
- To extract pitch-based features (pitch contour, velocity, acceleration)
- Improve accuracy in identifying raaga despite tonic variations and complex gamakas
- Create and expand the model so that it works well on both Melakartha and janya Raaga
- Efficiently classifies all carnatic music be it just Alapana, Kritis or a mixture of both. Currently focused on Krithis.

Dataset Description

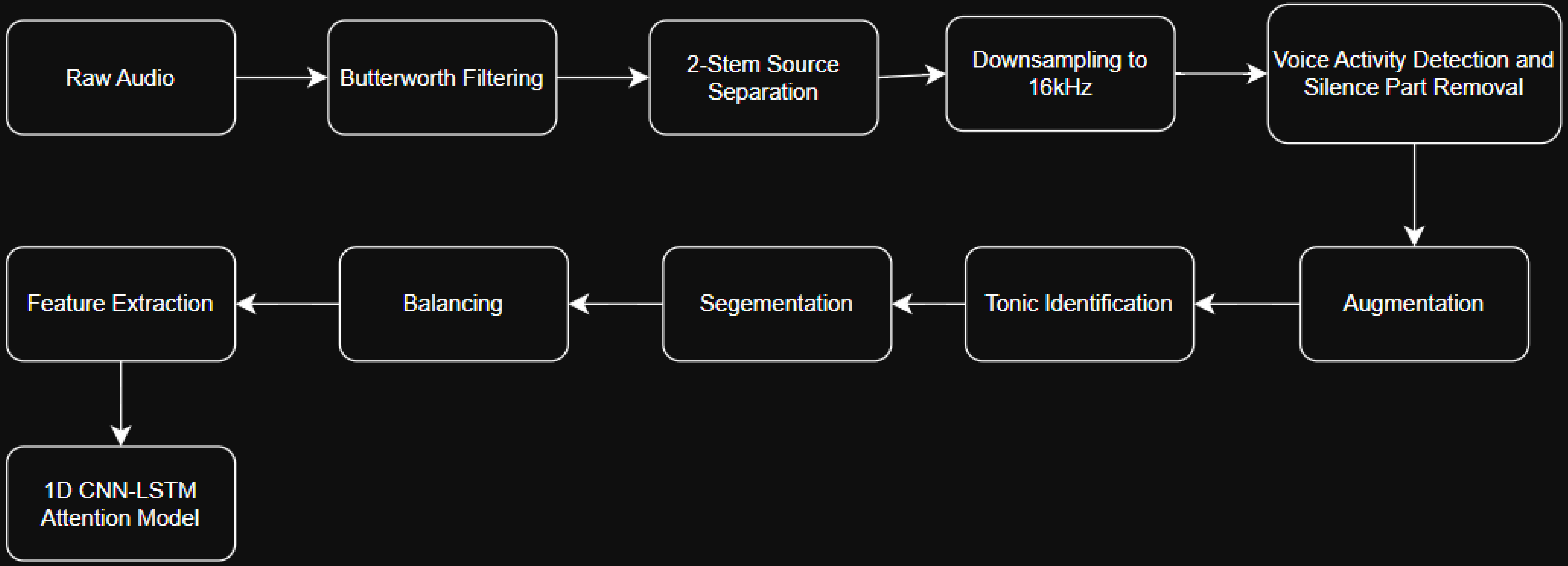
SwaraRaaga Sudha

Own Dataset

- Created own dataset by downloding audio tracks of each raaga from youtube.
- Consist of 10 raagas (mix of Melakarta and Janya Raaga)
- Each raaga have about 1 hour recordings
- Consist of different singers both male and female

Raaga Name	Total no. of Tracks	Hours	Female Artist	Male Artist
AnandaBhairavi	10	57 min 19 sec	6	4
Darbari Kannada	10	58 min 29 sec	4	6
Hamsadhwani	9	1 hour 1 min 5 sec	5	4
Kalyani	9	1 hour 1 min	2	7
Karaharpriya	7	1 hour 3 min 5 sec	3	4
Maayamalawagowla	7	1 hour 5 min	3	4
Mohanam	7	1 hour 3 min 50 sec	6	1
Neelambari	8	1 hour 28 sec	3	5
Shankarabharanam	6	1 hour 5 min 3 sec	4	2
Thodi	6	1 hour 4 min 54 sec	2	4

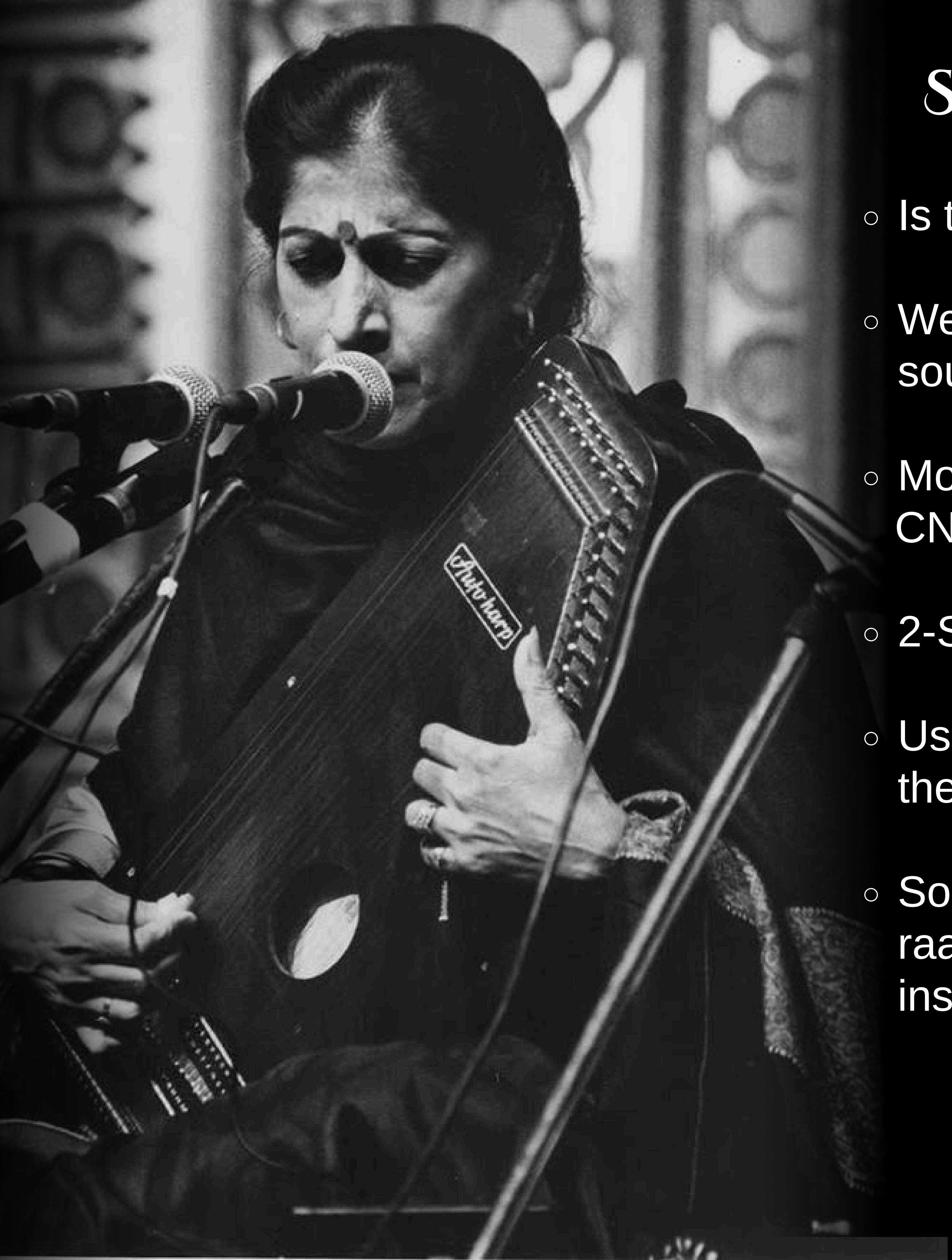
Methodology


Raw Audio
Butterworth Filtering
2-Stem Source Separation
Downsampling to 16kHz
Voice Activity Detection and Silence Part Removal
Augmentation
Tonic Identification
Segementation
Balancing
Feature Extraction
1D CNN-LSTM Attention Model

DATA PREPROCESSING

BUTTER WORTH FILTERING

- Applied a 5th-order Butterworth bandpass filter with a frequency range of 50 Hz to 5000 Hz.
- Vocal fundamental frequencies in Carnatic music typically lie between 80 Hz and 800 Hz.
- Expressive elements and ornamentalations (gamakas) can extend up to 5000 Hz.
- Frequencies below 80 Hz and above 5 kHz are often considered noise
- So,Butterworth filtering helps isolate musically relevant content and improves feature extraction quality.



Source Separation Using HTDemucs

- Is the process of isolating vocals and instruments
- We have used HTDemucs developed by Facebook AI Research for source separation
- Model contains a hybrid deep learning model combining CNN, Bi-LSTM and Transformer blocks.
- 2-Stem Source Separation is performed on the filtered files
- Usually HTDemucs will return all the sources but we have only saved the vocals.
- Source Separation ensures that certain intrinc characherters of the raaga obtained from vocals wont get mixed and get affected due to instrumets.

DOWNSAMPLING

- After source separation, our audio files were in 44.1 kHz format.
- WebRTC Voice Activity Detection (VAD) requires audio to be at a 8kHz, 16kHz, 32kHz or 48kHz sampling rate.
- To ensure compatibility with WebRTC along with making our further computational we downsampled the audio to 16 kHz.
- This was done using the librosa.resample() function.
- Downsampling ensures accurate and efficient silence detection using the VAD algorithm.

SILENCEREMOVAL USING WEBRTC VAD

- HTDemucs preserves the total duration of the audio files which leads to silent parts in the audio file.
- Along with this silence is present in places where the singer takes a pause.
- It is important to remove these as they are not usually important and can affect the model during training.
- For this we have used WebRTC VAD developed by Google.
- Audio files were split into 30 ms frames and labeled as voiced or unvoiced based on energy, spectral features, and temporal patterns.
- We used medium aggressiveness (level 2) to balance trimming silence and preserving low soft vocal phrases.
- At the end only the voiced frames were concatenated to reconstruct the audio without silence.

Augmentation

- We have used data augmentation for the purpose of increasing our dataset size and including variablility
- For this we have used the following data augmentation processes.

1. Pitch Shifting:
   - Shifted pitch by +2 semitones to simulate natural variations in vocal range.

2. Time Stretching:
   - Applied a stretch factor of 1.1 to alter the tempo slightly without affecting pitch.

3. White Noise Addition:
   - Added subtle background noise with noise factor 0.005 to improve robustness.

- Each audio clip was augmented in 3 different ways, increasing dataset size 4× (original + 3 variants).

TONIC ESTIMATION AND AUDIO SEGMENTATION

- Before segmenting the audio files, we have estimated the tonic frequency or Shadja (Swara Sa) using TorchCREPE.

- The tonic frequency varies among singers and it is important to tonic normalize the audio files for people indepentent raaga recognition.

- TorchCREPE is a GPU accelerated pitch estimation model.

- We have used a 10ms hope size for fine temporal resolution

- After Tonic frequency estimation the audio files were segmented into 10sec with 50 percent overlap

- Each segment was saved with its corresponding tonic value in the filename. E.g. Mohanam_tonic=131.2_seg0.wav In order to use for feature extraction

BALANCING

- To ensure unbiased training, it is important to have equal data representation across all raagas.
- For this we have done a duration based balancing where we iterate through all raaga files, finds the raaga with the least number of files and makes the rest of the raaga the same duration as the one with the least.
- In our dataset, Kharaharapriya had the least duration at 8920 seconds.
- For raagas with more data than this, only the first few segments (in sorted order) were selected to match the target duration.

FEATURE EXTRACTION

- In our work we have mainly focused on pitch based feature. including pitch contour, pitch velocity and pitch acceleration.

- For extracting these features we have used the CREPE model which is highly suitable for Indian classsical music because of its ability to detect continuous pitch glides.

- We used viterni decoding for smoother pitch trajectories and a 10ms step size for high temporal resolution.

Pitch Contour

- Represents how the fundamental frequency changes over time in a musical performance.
- It is further normalised into cent Scale for person-independent raaga classification.

$$P(t) = 1200 \cdot log_2 \left(\frac{f(t)}{f_0}\right)$$

Pitch Velocity

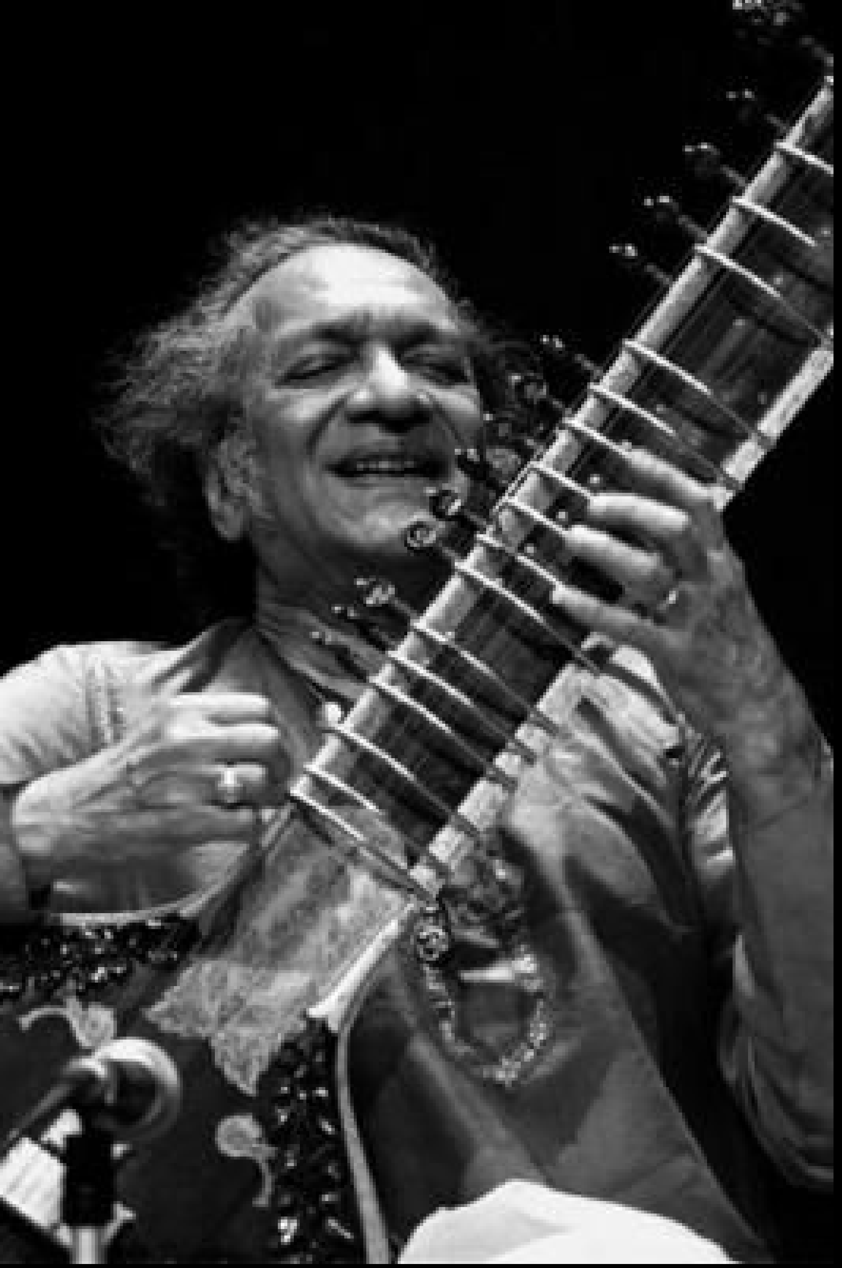
- Measures how quickly pitch changes over time.
- In Carnatic Music,the speed and the way in which transitions occurs define a raaga’s identity. For eg: Neelambari has slow, soothing oscillations, while Shankarabharanam has quick flicks.
- Itcan be found using the gradient ofpitch contour (in cents)

$$V(t) = \frac{dP(t)}{dt}$$

Pitch Acceleration

- Captures how quickly the rate of pitchchange increases or decreases.
- Tells us about the expression of gamakas.
- E.g., In Neelambari a phrase begin fast and then slow down smoothly.

$$A(t) = \frac{dV(t)}{dt} = \frac{d^2P(t)}{dt^2}$$



MODEL

DATASET SPILT

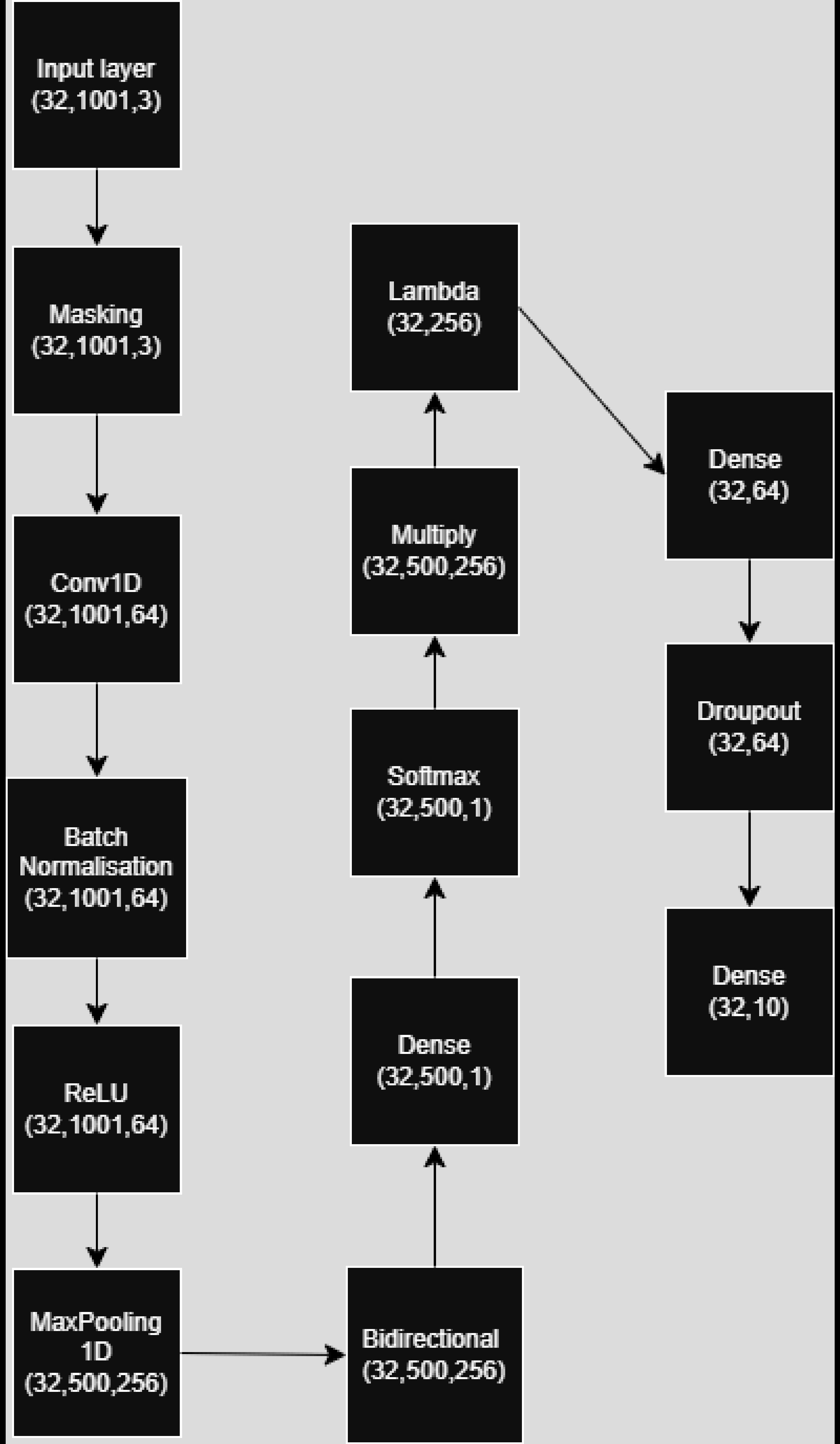
- Since we have used augmentation in our work, it was necessary to ensure that the testing and validation doesnt contain augmented data.
- In order to ensure this we have saved the segments with labels that would help in differentiating the augmented data with the ones.
- For augmented data the label contain words like ‘pitchshift’, ‘Noise’, and ‘stretch’.
- Based on this we have divided the data into 70 train, 15 val, and 15 test and saved them in 3 different folders

Train	21,264
Test	1,130
Val	904

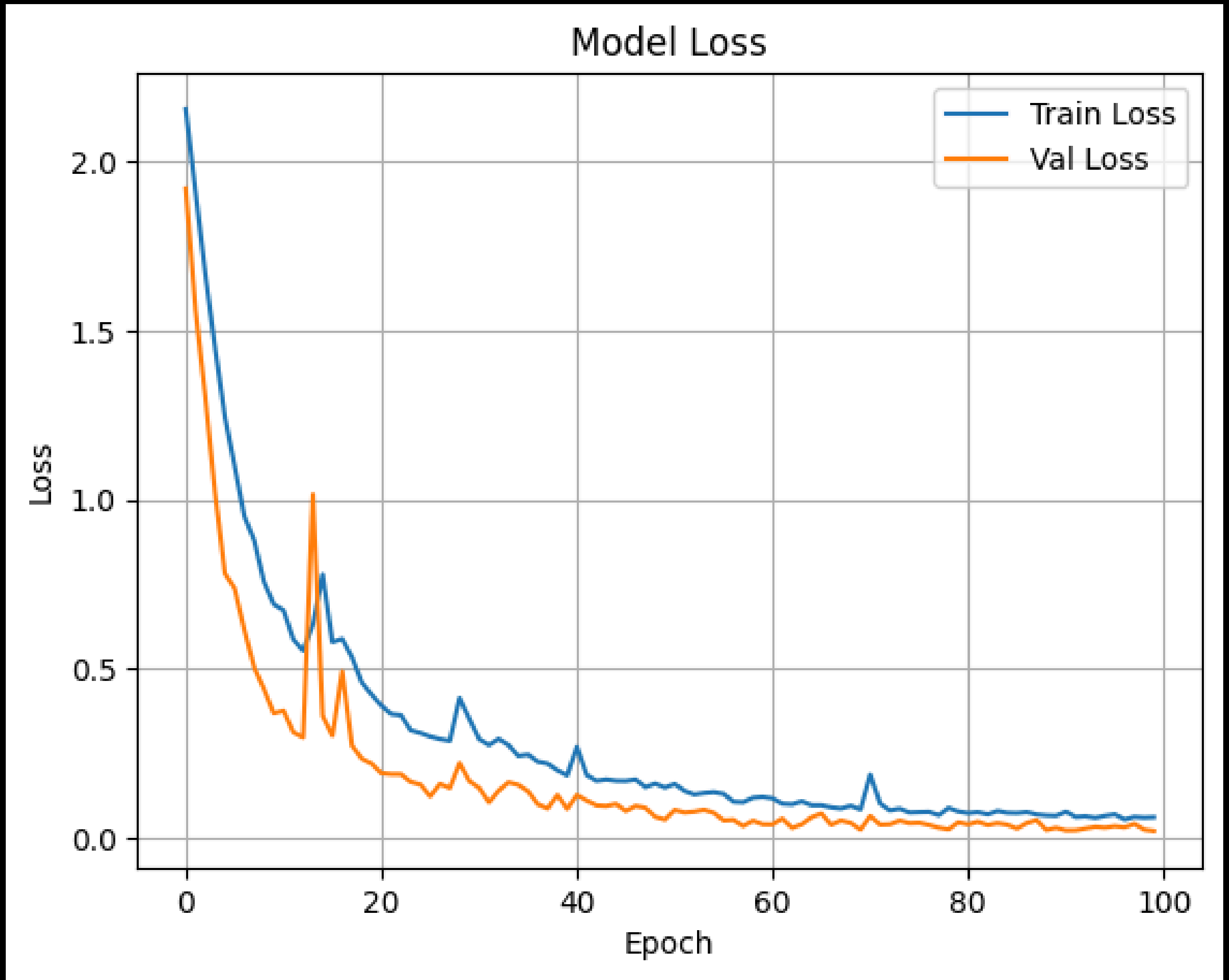
MODEL

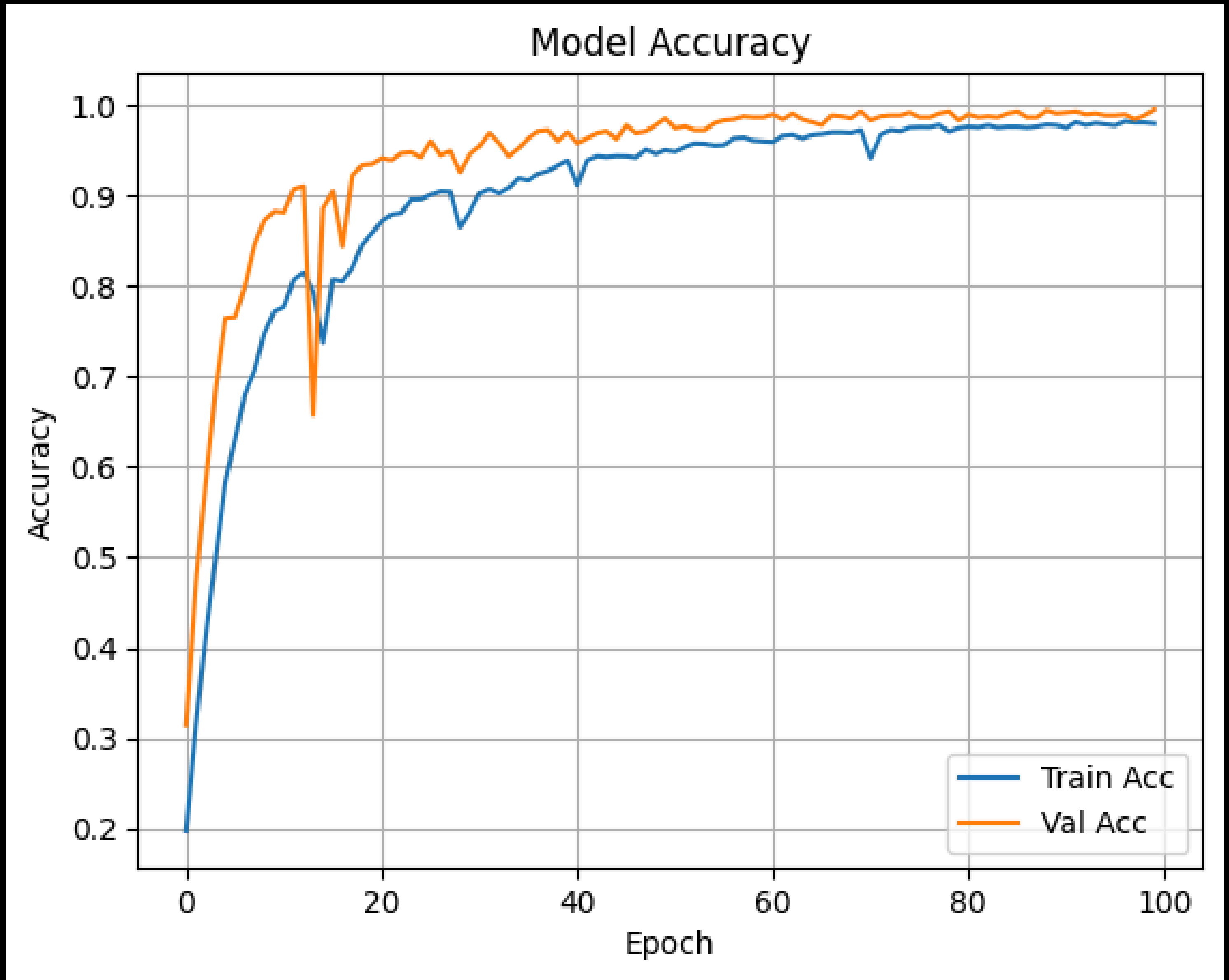
- For our work we have used a 1D CNN-LSTM with Atention model which is light weight.
- 1D CNN Detects local patterns in the timeseries ie, helps in capturing short-term variations in pitch, velocity and acceleration.
- Bidirectional LSTM learns long-term dependencies in sequential data from both past and future and helps in detecting patterns that depend on what came before and after.
- Attention Layer: Instead of treating every frame equally, the attention mechanism gives higher importance to time steps that contribute most to the classification ie the model learns to focus on the important time steps and less on irrelevantones.

Model Architectire



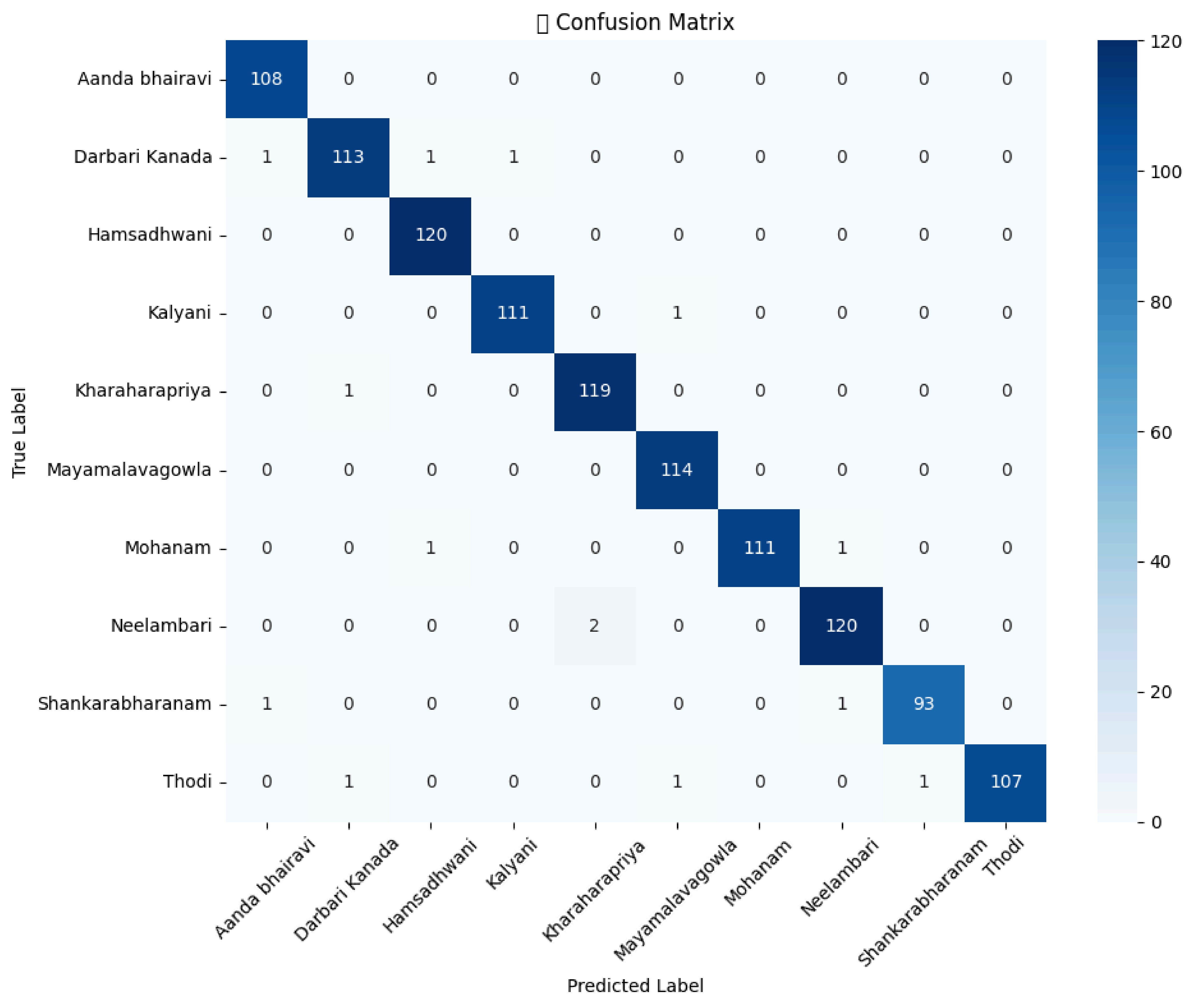
Results





Our model showed promising results with a testing accuracy of 98.76 and testing loss of 0.0451,when tested on completely unseen data.

Raaga	Precision	Recall	f1-Score	Support
Ananda Bhairavi	0.98	1.00	0.99	108
Darbari Kanada	0.98	0.97	0.98	116
Hamsadhwani	0.98	1.00	0.99	120
Kalyani	0.99	0.99	0.99	112
Kharaharapriya	0.98	0.99	0.99	120
Mayamalavagowla	0.98	1.00	0.99	114
Mohanam	1.00	0.98	0.99	113
Neelambari	0.98	0.98	0.98	122
Shankarabharanam	0.99	0.98	0.98	95
Thodi	1.00	0.97	0.99	110


Confusion Matrix
True Label
Predicted Label
Aanda bhairavi
Darbari Kanada
Hamsadhwani
Kalyani
Kharaharapriya
Mayamalavagowla
Mohanam
Neelambari
Shankarabharanam
Thodi
108 0 0 0 0 0 0 0 0 0
1 113 1 1 0 0 0 0 0 0
0 0 120 0 0 0 0 0 0 0
0 0 0 111 0 1 0 0 0 0
0 1 0 0 119 0 0 0 0 0
0 0 0 0 0 114 0 0 0 0
0 0 1 0 0 0 111 1 0 0
0 0 0 0 2 0 0 120 0 0
1 0 0 0 0 0 0 1 93 0
0 1 0 0 0 1 0 0 1 107
0
20
40
60
80
100
120

Conclusion

- In this work, we developed a deep learning-based approach to identify and classify Carnatic raagas from vocal recordings using a hybrid CNN-BiLSTM model with attention.

- The system was built upon a carefully designed preprocessing pipeline, including filtering, source separation, silence removal, tonic normalization, segmentation, and pitch-based feature extraction.

- By focusing on three expressive features ie, pitch contour, pitch velocity, and pitch acceleration we captured the dynamic melodic characteristics that define each raaga.

- The use of attention mechanisms helped the model focus on the most significant time segments within the audio, improving classification performance.

- Our model successfully demonstrates that Carnatic raagas can be computationally modeled and classified with high accuracy, contributing toward automated rāga recognition, music recommendation systems, and the preservation of Indian classical music through technology.

References

- S. John, S. R. S. Sudheesh, M. S. Sinith, and L. P. P., "Classification of Indian Classical Carnatic Music Based on Raga Using Deep Learning," in Proc. 2020 IEEE Recent Advances in Intelligent Computational Systems (RAICS), Trivandrum, India, Dec. 2020, pp. 110-113. doi: 10.1109/RAICS51191.2020.9332482.

- G. K. Koduri, S. Gulati, and P. Rao, "A Survey of Raaga Recognition Techniques and Improvements to the State-of-the-Art," in Proc. Sound and Music Computing Conf., Padova, Italy, 2011.

- D. P. Shah, N. M. Jagtap, P. T. Talekar, and K. Gawande, "Raga Recognition in Indian Classical Music Using Deep Learning," in Lecture Notes in Computer Science, vol. 12630, Apr. 2021, pp. 215-228. doi: 10.1007/978-3-030-72914-1_17.

- P. Chordia and A. Rae, "Raag Recognition Using Pitch-Class and Pitch-Class Dyad Distributions," in Proc. Sound and Music Computing Conf., Padova, Italy, 2007. Available: https://www.researchgate.net/publication/220723453

- A. K. Anitha, B. D. Parameshahcari, and Kiran, "Fisher-Jenks Classification Technique for Music Raaga Identification," in Proc. 2023 International Conference on Ambient Intelligence, Knowledge Informatics and Industrial Electronics (AIKIIE), 2023, pp. 1-7. doi: 10.1109/AIKIIR60097.2023.10390162.

References

- S. Paschalidou and I. Miliaresi, "Multimodal Deep Learning Architecture for Hindustani Raga Classification," Sensors & Transducers, vol. 261, no. 2, pp. 77-86, June 2023. [Online]. Available: http://www.sensorsportal.com.

- B. R. Bhagyalakshmi and M. B. Anandaraju, "Machine Learning Based Indian Raga Classification and Detection," in Proc. 2024 Asia Pacific Conference on Innovation in Technology (APCIT), Mysuru, India, Jul. 2024, pp. 1-5. doi: 10.1109/APCIT62007.2024.10673605.

- P. Singh and V. Arora, "Explainable Deep Learning Analysis for Raga Identification in Indian Art Music," J. LaTeX Class Files, vol. XX, no. XX, pp. XX-XX, 2024. [Online]. Available: arXiv:2406.02443.

- S. Patil, R. P. Bhavsar, and B. V. Pawar, "A Survey on Computational Approaches for Raga Identification," in Proc. 2022 IEEE World Conference on Applied Intelligence and Computing (AIC), 2022, pp. 85-92. doi: 10.1109/AIC55036.2022.9848984.

- T. Katte and B. S. Tiple, "Techniques for Indian Classical Raga Identification– A Survey," in Proc. 2014 IEEE INDICON Conference, Pune, India, Dec. 2014, pp. 1-6. doi: 10.1109/INDICON.2014.7030372