PROJECT ABSTRACT

Raga Recognition using CNN

A CNN classifier that identifies Hindustani ragas from audio, trained on a selected subset of the open Saraga dataset with log-mel and chroma features. Delivered built-to-order with training notebook, demo app, report, PPT and viva kit.

01. Title

Raga Recognition using CNN

02. Introduction / Background

A raga is far more than a scale: it is a melodic framework defined by characteristic phrases (pakad), ornamentation (meend, gamak) and note hierarchies that even trained listeners take years to internalize. Automating raga identification is a classic music-information-retrieval problem and an excellent deep-learning project, because the signal is genuinely hard — near-neighbour ragas share the same notes and differ only in how they move between them. The open Saraga dataset (CompMusic, Universitat Pompeu Fabra) provides annotated Hindustani recordings with raga labels. This project trains a compact 2D CNN on log-mel spectrograms and chroma features to classify 30-second clips into 10 ragas, served through a demo app with confidence-ranked predictions.

03. Problem Statement

Students and casual listeners have no accessible tool that names the raga of a performance they are hearing, and existing audio classifiers target Western genres, not Indian art music. Manual identification needs years of training, and naive pitch-histogram methods fail on ragas that share note sets. The project addresses this by building a CNN that learns time-frequency melodic shapes from real Hindustani recordings: tonic-normalized chroma and log-mel inputs let the network see ornamentation and phrase movement, not just static pitch content, and the training notebook reports accuracy, macro-F1 and the full confusion matrix honestly.

04. Objectives

- Curate a 10-raga subset (ragas with the most recordings) from the Saraga Hindustani collection (108 recordings, 43.6 hours, 61 unique ragas).
- Build a preprocessing pipeline: 30-second clips, tonic normalization, 128-band log-mel spectrograms and 12-bin chroma features.

- Train a compact 2D CNN (~1.2M parameters) with SpecAugment-style augmentation, targeting approximately 80% clip-level accuracy on held-out recordings.
- Build a demo app: upload/select a clip, view its chromagram, get top-5 raga predictions with confidence bars.
- Deliver the complete student kit: training notebook, demo app, report, PPT and viva Q&A.

05. Proposed Methodology

The build has three stages. (1) Data pipeline: Saraga Hindustani recordings are segmented into 30-second clips; splits are made at the recording level so no clip from one recording leaks between train/validation/test. Each clip is converted to a 128-band log-mel spectrogram plus tonic-normalized 12-bin chroma. (2) Model: a 2D CNN with three convolutional blocks (3x3 conv, batch norm, ReLU, max-pool), two dilated conv blocks with dropout, time-axis global average pooling, a 256-unit dense layer and a 10-way softmax — trained with Adam, cosine decay and early stopping for about 40 epochs. (3) Evaluation and demo: the notebook logs accuracy, macro-F1, per-class precision/recall and the confusion matrix; the demo app wraps inference with chromagram visualization and confidence-ranked predictions.

06. System Architecture / Working

At training time, clips flow through augmentation (time/frequency masking, +-1 semitone pitch shift), the CNN, and a cross-entropy loss, with metrics logged every epoch. At inference time the demo app takes an audio clip, computes the same log-mel + chroma features, runs the trained weights, and displays the top-5 raga probabilities with a chroma heatmap showing the pitch-class energy the network attended to. The confusion matrix from the notebook documents exactly which raga pairs the model confuses — expected to be near-neighbour ragas sharing note sets — which becomes the honest core of the report's error analysis.

07. Hardware / Software Requirements

- Python 3.10, PyTorch (CNN training and inference)
- librosa (audio loading, mel spectrograms, chroma), NumPy, scikit-learn (metrics)
- Matplotlib, Seaborn (training curves, confusion matrix)
- Saraga Hindustani collection subset (open dataset, CompMusic)
- Flask or static demo app for the prediction interface
- CPU training is feasible for the compact model; a GPU speeds up experimentation

08. Expected Outcomes

- Trained CNN weights for 10-raga classification from 30-second clips.

- Training notebook logging accuracy, macro-F1 and confusion matrix on held-out recordings.
- Demo app with clip selection, chromagram visualization and top-5 predictions.
- Design target of approximately 80% clip-level accuracy (target, not a measured claim).
- Complete documentation: report, PPT, dataset guide and viva Q&A.

09. Applications

- Educational tool for students learning Hindustani raga identification.
- Teaching material for audio deep learning: spectrograms, chroma, CNNs, augmentation.
- Reference build for other Indian-music MIR tasks (tala, instrument recognition).
- Concert companion prototype showing likely raga during a performance.

10. Future Scope

- Full 61-raga classification on the complete Saraga Hindustani collection.
- Attention-based models (transformers) over longer temporal contexts.
- Real-time raga tracking on live audio streams.
- Cross-collection evaluation on Carnatic recordings.

11. References

- Srinivasamurthy, A. et al. — Saraga: Open Datasets for Research on Indian Art Music (Empirical Musicology Review); CompMusic, Universitat Pompeu Fabra.
- Salamon, J. and Bello, J. P. — Deep Convolutional Neural Networks and Data Augmentation for Environmental Sound Classification (2017).
- Choi, K. et al. — Automatic Tagging Using Deep Convolutional Neural Networks (ISMIR 2016).
- librosa documentation — audio analysis library (librosa.org).