International Journal of Applied Mathematics

Vol 38 No 5,2025
ISSN: 1311-1728(printed version); ISSN: 1314-8060(on-line version)

A NOVEL APPROACH FOR SINGER IDENTIFICATION AND VOCAL RANGE ESTIMATION USING HYBRID DEEP LEARNING MODEL

Ms. Renju K [1], Dr. Ashok Immanuel V [2]

[1]Research Scholar, CHRIST (Deemed to be University), Bengaluru, Karnataka, India

[1]Assistant Professor, Mount Carmel College, Autonomous, Bengaluru, Karnataka, India
[1]renju.k@mccblr.edu.in

[2]Professor, CHRIST (Deemed to be University), Bengaluru, Karnataka, India
[2]ashok.immanuel@christuniversity.in

ABSTRACT

Automatic singer identification and estimating vocal range of singer are essential in music information retrieval, with significant applications in music education, digital archiving, personalized training, vocal health monitoring and assisting composers in selecting suitable singers for specific compositions. While previous research focused on singer classification using timbre and spectral features, computational approaches for vocal range estimation of singer remain underexplored. Identification of singer and analysis of vocal range will have a great impact on commercial and academic domains. In this research, we propose a hybrid PA (Pitch Analysis) model that employs CREPE deep learning model for pitch extraction, capturing pitch characteristics directly from vocal tracks after isolating it from instruments in a polyphonic recording. The extracted pitch values are then transformed into scalar embeddings and provided as an input to a SRT (Singer Range Transformer) model, which is used to train a Transformer encoder architecture to identify singers and extract their highest and lowest range of pitch. A customized dataset was curated to ensure diversity and robustness, consisting of twenty distinct Carnatic music compositions by the same singer. Experimental results demonstrated that the proposed model achieved high accuracy in identifying the singer and provides reliable vocal range estimations consistent with musicological expectations. This research introduces a novel computational approach to singer identification and vocal range analysis, offering valuable contributions to music education, training, archiving and vocal-health wellness.

Keywords: Singer Identification, Vocal Range Estimation, Carnatic Music, Pitch Analysis, Deep Learning, Transformer, CREPE

I. INTRODUCTION

Music Information Retrieval (MIR) has emerged as a key area of research enabling new avenues in music analysis especially with the advancement of machine learning and deep learning techniques. With the rise of large-scale digital music archives, singer identification and vocal range estimation becomes prominent among the various other tasks in MIR due to their applications in music education, performance analysis, digital preservation, and music production. Automatic singer identification helps systems recognize and classify singers based on their unique vocal pitch characteristics. This supports tasks such as digital archiving, generating metadata, and creating recommendation systems. The information regarding vocal range estimation help the singers to constantly monitor vocal health and take

Received August 11,2025

Vol 38 No 5,2025
ISSN: 1311-1728(printed version); ISSN: 1314-8060(on-line version)

necessary steps to improve their vocal range.Vocal range estimation is essential as it provide insights into a singer's lowest and highest attainable pitches and also helps us to understand how the age factor affect the vocal range of a singer. The male and female vocal ranges are different and this knowledge would provide the researchers to understand the gender based physiological differences in singing and speaking. It also helps in categorizing singers across various music genres. Although researchers have made remarkable impact on singer identification, most of the existing methods primarily focus on timbre and features from audio. These approaches showed promising results for classification but often ignore pitch dynamics. This is particularly relevant in music especially in Indian Carnatic music where melodic structure and range shape identity of the singer.

The analysis of vocal range of a singer has several practical applications that extend beyond academic research. The information regarding a singer's vocal range is precious for music directors and composers when new songs are assigned to the singers ensuring that the chosen compositions fall within the singer's comfortable pitch range. This not only enhances the quality of the performance but also reduces the recording time and effort. Singers themselves can take the advantage from range estimation as it provides insights into their strengths and weakness, allowing them to undertake targeted vocal training exercises to expand their range or improve control over specific registers. The precise vocal range estimation allows compositions to be transposed into keys that best suit the singer, resulting in optimized performances free of vocal strain in live concerts and studio recordings. Regular monitoring of vocal range can also serve as an early diagnostic tool for vocal health, as reductions in range may indicate fatigue, overuse, or medical concerns that require attention. Singers can be benefited with vocal range data which can support personalized music recommendation systems, such as karaoke applications that suggest songs within a user's range. Furthermore, in musicological studies, systematic range estimation provides deeper insights into stylistic differences between singers, traditions, or genres, offering invaluable contributions to the study of Indian Carnatic and contemporary music.Moreover, there has been limited research on estimating vocal range within a computational framework, leaving a significant gap in MIR studies. Addressing this gap would benefit both practitioners and learners by providing objective, data-driven assessments of vocal abilities.

In our study, we proposes a hybrid PA (Pitch Analysis) model that uses deep learning techniques for automatically identifying singer and hence estimating singer's vocal range. The proposed model utilizes the CREPE(Convolutional Representation for Pitch Estimation) model for pitch extraction, transform pitch values into scalar embeddings which are then processed by a SRT(Singer Ran.ge Transformer), Transformer encoder-based recognition model. The model, CREPE is a state-of-the-art deep learning model designed for monophonic pitch detection [1]. Unlike traditional pitch-tracking algorithms like PYIN and PYAAPT that rely on hand-crafted features such as autocorrelation, cepstrum, or harmonic summation, CREPE directly operates on raw time-domain audio waveforms [2][3]. CREPE model capture both fine-grained temporal structures and harmonic patterns with a deep convolutional neural network having six convolutional layers to learn hierarchical representations of audio [4]. By training these pitch values on a custom Carnatic music dataset that includes twenty different performances by the same singer, the model achieves high accuracy in recognizing singer and hence extract the highest and lowest range predictions that align with musicological expectations. This research contributes an innovative methodology for MIR and thereby improves computational methods for identifying the singer and estimation of vocal range.

II. LITERATURE REVIEW

Serhat Hızlısoy et.al implemented a study on identifying singer from a dataset of Turkish songs[5]. The proposed model leverages acoustic features such as mel-spectrograms and octave-based spectral contrast values to construct a hybrid feature vector for classification. The approach addressed challenges such as variations across albums, years, and singing styles, showing stable performance in recognizing singers. The dataset comprised 180 songs from 9 Turkish singers, with feature reduction via PCA, normalization, and classification using Extra Trees and Logistic Regression, achieving an accuracy of 89.4%, with precision, recall, and F-score around 89–90%. Additionally, evaluation on the widely used artist20 dataset yielded 85.4% accuracy with KNN, demonstrating the model's applicability beyond Turkish songs. The findings highlight the significance of combining spectral features with

Received August 11,2025

Vol 38 No 5,2025
ISSN: 1311-1728(printed version); ISSN: 1314-8060(on-line version)

dimensionality reduction and normalization to improve singer identification. The authors also suggest extending the model to multilingual datasets and multi-singer recognition tasks in the future.

Murthy et.al investigated both conventional feature-based methods and CNN-based techniques for identifying singers in the context of Music Information Retrieval [6]. The study utilized two datasets: artist20 and a database of popular Indian singers featuring 20 artists. The features extracted were cepstral features (MFCCs, LPCCs), temporal features (SDC), and chroma features, creating a 46-dimensional feature vector. A genetic algorithm-based feature selection (GAFS) method was employed for feature selection, and classification was performed using ANN and Random Forest. Furthermore, spectrograms and chromagrams were directly fed into CNNs. The findings indicate that CNNs outperformed traditional classifiers, achieving approximately 75% accuracy on the Indian singers' database, while the performance on the artist20 dataset remains lower, not surpassing 60%.

Abhale et.al proposed a neural network based on Long Short-Term Memory (LSTM) that is paired with vocal separation methods to extract vocal segments and classify singers [7]. The system utilized Python library Librosa for feature extraction from audio files, applied spectrogram analysis to identify vocal components, and trained a LSTM model to detect singer-specific patterns. By employing Mel-Frequency Cepstral Coefficients (MFCCs) and chroma features, the solution provided a robust audio representation, resulting in high classification accuracy. The paper elaborated on the system's architecture, implementation challenges, experimental outcomes, and future improvements, such as real-time processing and multi-singer detection. Theproposed LSTM-based framework effectively tackled the challenge of singer identification in polyphonic music by combining vocal separation, temporal feature extraction, and deep learning. The highlights of this research paper includes the use of harmonic-percussive source separation (HPSS) to isolate vocal segments and the application of bidirectional LSTM layers to model long-term dependencies in MFCC and chroma features. Experimental results on the MIRIK dataset show a classification accuracy of 92.3%, surpassing traditional methods like SVM (85.2%) and MLP (88.1%). This underscores the superiority of LSTM networks in capturing dynamic vocal patterns and reducing noise interference.For singer identification, Shichao Hu et.al examined various strategies within a deep metric learning framework, focusing on their effectiveness with a large-scale dataset containing audio samples from 5057 singers [8]. The research implemented comprehensive experiments to evaluate various loss functions, such as triplet loss, generalized end-to-end (GE2E) loss, and prototypical network (PN) loss. The impact of vocal source separation were also analyzed in this paper. Their findings were, when using audio inputs with isolated vocals, model trained with PN loss surpasses other methods in the identification task. In contrast, for the verification task involving a one-on-one comparison of two single embeddings, triplet loss yields the best outcomes. However, when using the centroid of 5 embeddings to represent the singer embedding, verification with PN loss outperforms methods using triplet loss. The research also revealed the fact that employing longer segments for representing a singer consistently enhances performance across all evaluated tasks.

Deepali Y et.al explored a Singer Identification System(SID) tailored for Indian playback singers by examining a vast dataset [9]. The model extracted four distinct acoustic features from singing voice segments such as formants, harmonic spectral envelope, vibrato, and timbre which uniquely identifies the singer. By integrating these multiple acoustic features, the study addressed major SID challenges such as variations in a singer's voice, testing multilingual songs, and the album effect. Systematic evaluation demonstrated that the SID is resilient to variations in singing style and song structure and is effective in identifying cover songs and singers. The results were based on an in-house cappella database comprising 26 singers and 550 songs. By reducing the dimensionality of the feature vector and employing a Support Vector Machine classifier, an accuracy of 86% was achieved using a fourfold cross-validation process. Furthermore, a performance comparison with other existing methods highlights the superiority of the proposed approach in terms of dataset volume and song duration. This research proposed a framework that leads to robust singer identification by considering multiple acoustic features of the singing voice. The feasibility of the system was tested on a database of pure cappella samples from Bollywood playback singers, and the performance of these acoustic features in capturing singer characteristics was assessed. The experimental findings indicate that utilizing multiple features can effectively capture singer characteristics from a diverse database.

Received August 11,2025

International Journal of Applied Mathematics

Vol 38 No 5,2025
ISSN: 1311-1728(printed version); ISSN: 1314-8060(on-line version)

A self-supervised learning (SSL) framework was proposed by Bernardo Torres et.al for training singer identity encoders, aiming to extract robust voice representations applicable to tasks such as singer similarity and singing synthesis.[10] A large corpus of isolated audio tracks were used with data augmentations to achieve invariance to pitch and lyrical content. The models were evaluated across multiple datasets for singer identification and similarity, with particular attention to out-of-domain generalization. The results showed that the SSL models outperformed speaker verification and wav2vec 2.0 baselines, producing high-quality embeddings at 44.1 kHz without requiring specialized architectures. Among methods tested, BYOL generalized best to out-of-domain data, while contrastive approaches worked better in-domain. Compared to XLSR-53, performance was slightly lower, but the models surpassed wav2vec-base with fewer parameters. Limitations remain in handling unique singing techniques (e.g., VocalSet), highlighting the need for more robust SSL frameworks. Overall, the findings demonstrated that self-supervised methods hold strong potential for advancing singer identification and similarity tasks.

This study addressed the challenges of automatic singer identification (SID) in popular music recordings, where background accompaniment often interferes with the clarity of the singer's voice[11]. The authors proposed KNN-Net, a deep neural network designed to learn local timbre feature representations from polyphonic audio signals. Unlike conventional models that rely on a softmax output layer, KNN-Net incorporated a KNN-based output layer for improved interpretability, along with an attention mechanism to emphasize critical timbre features. The experimental evaluation on the artist20 dataset demonstrated a 4% improvement over state-of-the-art methods, while additional testing on newly created singer32 and singer60 Chinese pop music datasets confirmed the robustness and effectiveness of the model. Overall, the results highlighted the potential of KNN-Net and attention mechanisms in enhancing singer identification task.

Pham Van Toan et.al emphasizes the importance of singer voice classification as identifying singers plays a crucial role in music information retrieval, content indexing, and organization of large music libraries [12]. Their study introduced a method for singer identification in Vietnamese popular music, where vocal segment detection and singing voice separation are employed as preprocessing steps to isolate the singer's voice from background accompaniments. For classification, a neural network model was trained using MFCC features extracted from the separated vocals. The approach was evaluated on a dataset of 300 songs by 18 well-known Vietnamese singers, achieving an accuracy of 92.84% with 5-fold stratified cross-validation, outperforming existing methods on the same dataset. The framework integrated three stages—vocal segmentation, vocal extraction, and classification—each optimized with suitable neural network architectures. With an overall accuracy of about 93%, this work demonstrated the effectiveness of deep learning for singer identification and provides a publicly available dataset to support further research in this area.

III. METHODOLOGY

The methodology implemented in this research is shown in Fig 1. While previous research in singer identification has relied on a wide range of acoustic and spectral features, our study focuses exclusively on pitch characteristics, since in Carnatic music the same pitch framework is consistently employed across all performances of a singer.

A. About Dataset

The dataset curated for this study consists of vocal recordings from two professional Carnatic singers, with twenty different performances per singer, resulting in a total of 40 recordings. The performances include kritis and other Carnatic compositions rendered across diverse ragas, thereby ensuring variability in melodic and rhythmic content. The inclusion of both male and female voices ensures diversity in vocal timbre and pitch range, which is particularly important for the task of vocal range estimation, as male singers typically occupy lower registers while female singers span higher pitch regions.

Received August 11,2025

International Journal of Applied Mathematics

Vol 38 No 5,2025
ISSN: 1311-1728(printed version); ISSN: 1314-8060(on-line version)

B. Preprocessing

In this research, Librosa library in Python was used for audio analysis. All recordings were resampled to 22,050 Hz, set the duration to 120 seconds and converted to mono for consistency. To avoid interference from accompanying instruments, the Spleeter tool was employed to isolate the vocal track. Pitch values were extracted using the CREPE algorithm, which provides frame wise pitch level tracking and confidence estimates. Unvoiced frames and low-confidence outputs were discarded, and the remaining pitch sequences were segmented into fixed-length inputs. Shorter sequences were padded to maintain uniform dimensionality, enabling compatibility with the Transformer Encoder model. For training, each singer was assigned a unique label that was converted into integer values through label encoding. To address dataset imbalance, class balancing techniques were applied so that each singer contributed an equal number of training sequences, ensuring that the model did not bias towards a particular class.

C. Audio Source Separation

In order to have more precise and reliable analysis on the vocals, Spleeter tool was used to separate vocals from Carnatic music recordings. The direct analysis of audio can lead to interference of instruments and hence may not be able to achieve high accuracy. Spleeter tool enables the separation of vocal track and hence reducing the overlapping of accompanying instruments by decomposing the audio into four stems- vocals, drums, bass and others, facilitating more detailed analysis.

D. PA (Pitch Analysis) Model

An innovative hybrid PA model was developed, which integrates the CREPE model to extract the pitch characteristics from the vocals isolated using the spleeter tool. The model outputs frame-wise pitch estimates along with a confidence score for each prediction, indicating the reliability of the estimate. For this research, only pitch values with a confidence level greater than 95% were considered. This threshold was chosen to ensure high reliability and accuracy of the extracted features, as lower-confidence estimates are more susceptible to noise. By filtering out uncertain pitch estimates, the model minimizes errors in pitch tracking and enhances the robustness of vocal range estimation. The scalar pitch values of various performances by the same singer obtained from CREPE with confidence greater than 95% are stored in a csv file used for training the SRT model.

E. SRT (Singer Range Transformer) Model

The hybrid PA model integrates SRT model, which is used to identify the singer and estimate the vocal range of the singer which is the main highlight of this research. The scalar pitch values obtained from PA model are converted into pitch embeddings through an embedding layer. The pitch embeddings, is a vector representation which carries not only the raw frequency numerical values but also includes additional musical information such as pitch class, tonic alignment, and interval relations. This representation transforms the numerical pitch sequence into a musically meaningful feature space suitable for deep sequence modelling.

The pitch embeddings are then passed into a SRT model which incorporates Transformer Encoder architecture. The main advantage of SRT model is that, the model incorporates self attention mechanism which looks at all time steps at once unlike other recurrent models which process the pitch values in sequence and may not perform well with long term dependencies. For instance, if the singer is singing a high note only at the beginning of a song, attention mechanism will treat this rare note as significant, instead of treating all notes as equal or ignoring the rare note. This information is very crucial in estimating the vocal range of the singer where it points to the singers ability to sing in high range. The resulting representations are then transferred to the last layers of network, after the attention mechanism projects the musically important pitch events.

Received August 11,2025

International Journal of Applied Mathematics

Vol 38 No 5,2025
ISSN: 1311-1728(printed version); ISSN: 1314-8060(on-line version)

As the model has been trained on pitch sequences of multiple singers, it learns to associate unique pitch patterns with specific singers. The model gives predictions for each segment of the pitch sequence for singer identification and to attain a final decision, a majority voting scheme is applied across all predictions at the frame level. The identity of the singer is established with largest number of votes from the frame-level predictions and this ensures that the final prediction reflects the pitching style and individuality of singer. The vocal range of the singer is estimated after revealing the identity of the singer by considering the highest and lowest range of the singer.

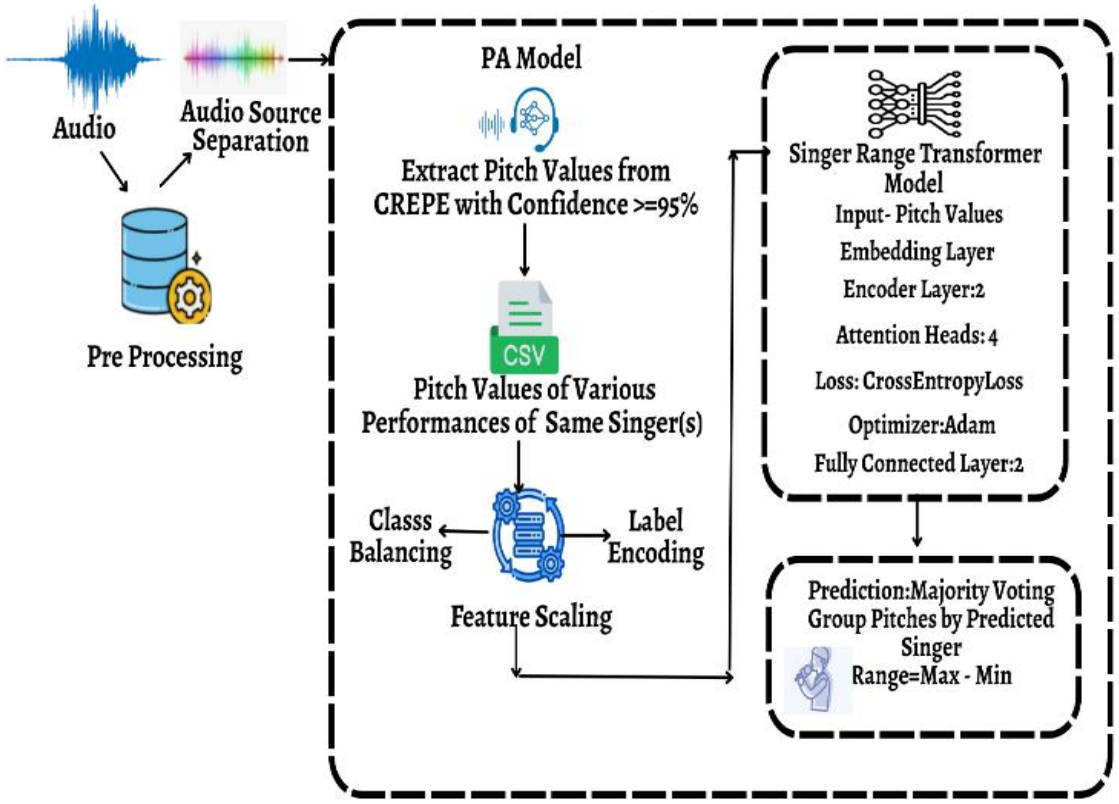

Fig 1: Hybrid PA Model integrated with SRT Model for Singer Identification and Estimation of Vocal Range

Received August 11,2025

International Journal of Applied Mathematics

Vol 38 No 5,2025
ISSN: 1311-1728(printed version); ISSN: 1314-8060(on-line version)

IV. RESULTS AND DISCUSSION

This study examined the role of pitch characteristics in the identification of singer and estimation of vocal range of very popular carnatic musicians Bombay Jayashree and Yesudas. The pitch histogram and boxplot representation of singers is shown in Fig 2 and Fig 3 respectively, provides additional insight into the classification behavior.

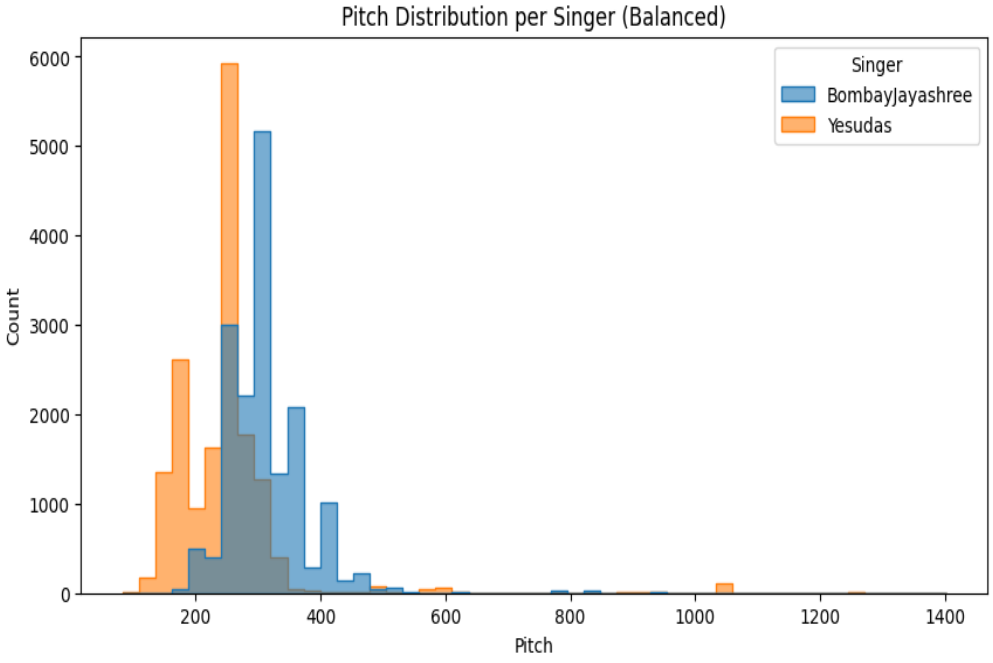

Fig 2: Histogram Representation of Pitch Distribution of Singers

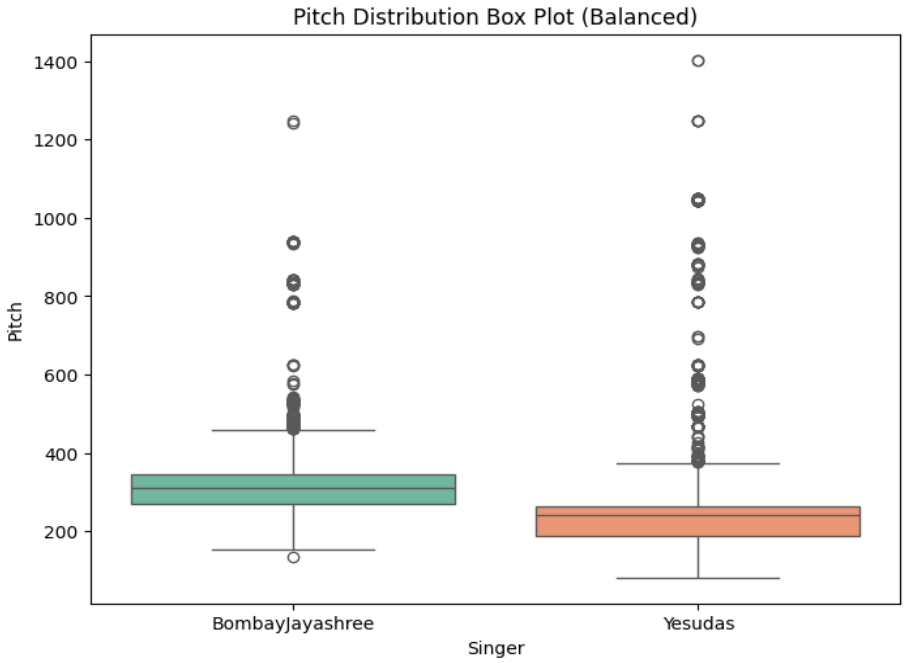

Fig 3:Boxplot Representation of Pitch Distribution of Singers

Received August 11,2025

International Journal of Applied Mathematics

Vol 38 No 5,2025
ISSN: 1311-1728(printed version); ISSN: 1314-8060(on-line version)

The pitch values of Yesudas is concentrated more in the lower-to-mid frequency range, while Bombay Jayashree's pitch values cluster in the mid-to-higher frequency regions. However,the two distributions show overlap, particularly in the 200–400 Hz range, which are shown in the observed misclassifications in the confusion matrix Fig 4.The model was evaluated on multiple analyses including classification performance, pitch distribution, learning behavior, and prediction confidence. The confusion matrix Fig 4 demonstrates the classification outcomes for the two singers.

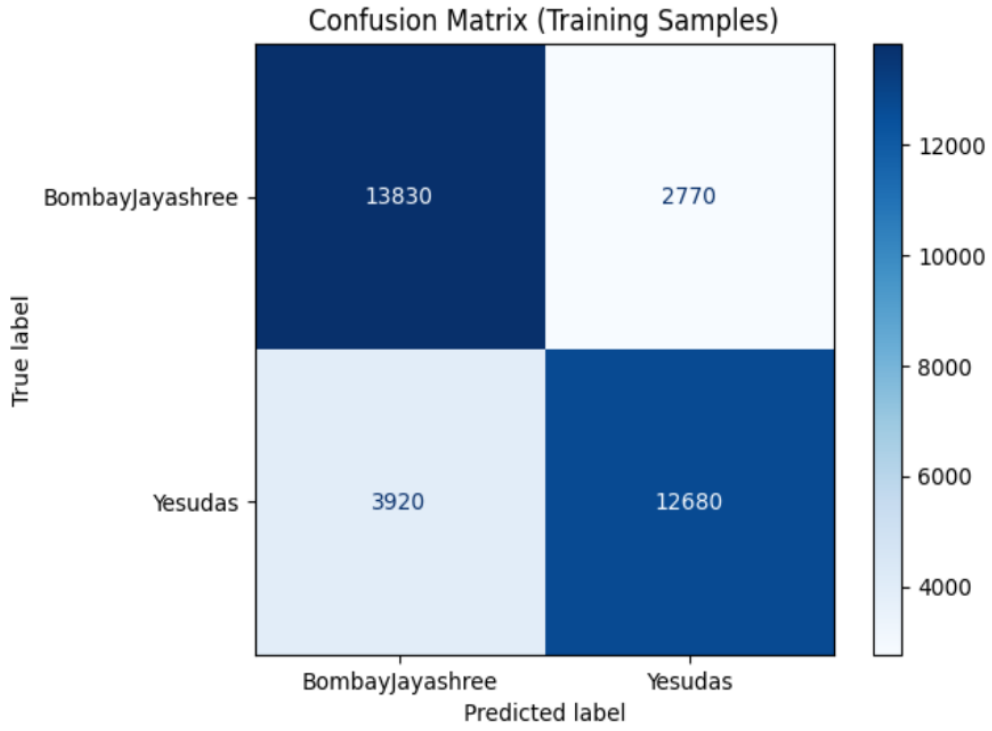

Fig 4: Confusion Matrix for Classifying Singers-Bombayjayashree & Yesudas

A total of 13,830 samples were correctly classified for the singer Bombay Jayashree, while 2,770 were misclassified as Yesudas. Similarly, for the singer Yesudas, a total of 12,680 were correctly classified and a total of 3,920 samples were misclassified. The model captured some discriminative pitch patterns, there remains significant overlap indicating the fact that both singers share same pitch space. This pitch information is sufficient as the singers would choose same pitch across multiple performances especially in Carnatic music.

After identifying the singer, the vocal range is calculated by finding the lowest(min) and highest (max) pitch of the predicted singer. The Fig 5 shows that the predicted singer is "Yesudas" and the lowest pitch is 123.52 Hz and highest pitch is 257.51Hz.The vocal range is then computed by taking the difference between highest and lowest range.

Received August 11,2025

Vol 38 No 5,2025
ISSN: 1311-1728(printed version); ISSN: 1314-8060(on-line version)

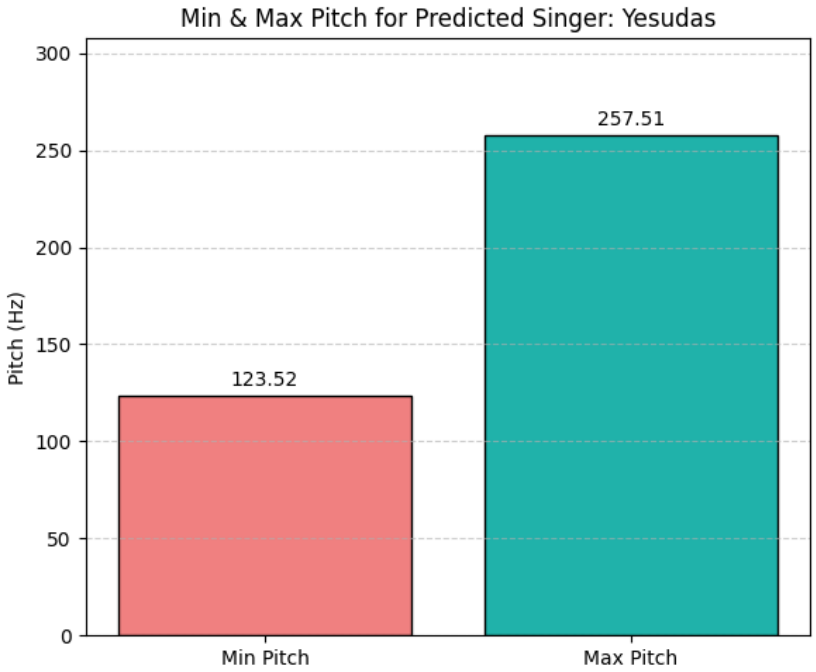

Fig 5: Vocal Range Estimation of the Predicted Singer

VI. CONCLUSION

This study proved an innovative method for singer identification and vocal range estimation based on a hybrid Pitch Analysis (PA) and Singer Range Transformer (SRT) model. The model was trained on various Carnatic music performances of the same singer to ensure that the system captured the inherent pitch characteristics peculiar to each singer. In Carnatic music, musicians render all compositions at their comfortable pitch, making the approach more reliable and robust. Having pitch outputs available at fixed 10-second intervals provided a richer dataset and made analysis both comprehensive and efficient.One of the main highlight of this research is the novel estimation of singers' vocal ranges, not broadly tackled in previous studies. Although the current implementation was restricted to two singers, the approach has good potential to be applied to a number of singers and other musical areas. Using this technique on other musical genres, such as Western classical, Hindustani, or popular music, would be useful to consider its applicability across the musical traditions. The findings validate that singer identification by pitch-based modeling is not merely successful but also introduces new possibilities for computational investigation of song capabilities in music.

REFERENCES

[1] Xavier Riley, Simon Dixon, CREPE Notes: A new method for segmenting pitch contours into discrete notes, Proceedings of the Sound and Music Computing Conference 2023, Stockholm, Sweden.
[2] Anja Kroon, Comparing Conventional Pitch Detection Algorithms with a Neural Network Approach, arXiv:2206.14357v1, June 2022
[3] Behnam Faghih, Joseph Timoney, An investigation into several pitch detection algorithms for singing phrases analysis, IEEE, https://doi.org/10.1109/ISSC.2019.8904943, June 2019
[4] Jong Wook Kim, Justin Salamon, Peter Li, Juan Pablo Bello, Crepe: A Convolutional Representation For Pitch Estimation, arXiv:1802.06182v1 [eess.AS] 17 Feb 2018

Received August 11,2025

Vol 38 No 5,2025
ISSN: 1311-1728(printed version); ISSN: 1314-8060(on-line version)

[5] Abhale B A , Dr. Rokade P.P, Sanap Abhishek, Kasar Pushkaraj, Thorat Pranjal ,Gudaghe Sachin, International Journal of Research Publication and Reviews, Vol 6, Issue 6, June 2025, ISSN 2582-7421

[6] Serhat Hızlısoy, Recep Sinan Arslan, Emel Çolakoğlu, Singer identification model using data augmentation and enhanced feature conversion with hybrid feature vector and machine learning, EURASIP Journal on Audio Speech and Music Processing, February 2024, DOI:10.1186/s13636-024-00336-8

[7] Y. V. Srinivasa Murthy, Shashidhar G. Koolagudi, T. K. Jeshventh Raja, Singer identification for Indian singers using convolutional neural networks, International Journal of Speech Technology, Volume 24, Issue 3,September 2021, https://doi.org/10.1007/s10772-021-09849-5

[8] Shichao Hu, Beici Liang, Zhouxuan Chen, Xiao Lu, Ethan Zhao, Simon Lul, Large-scale singer recognition using deep metric learning: an experimental study,2021 International Conference on Neural Networks, IEEE, September 2021, https://doi.org/10.1109/IJCNN52387.2021.9533911

[9] Deepali Y. Loni, Shaila Subbaraman, Robust singer identification of Indian playback singers, EURASIP Journal on Audio Speech and Music Processing, Volume 2019

[10] Bernardo Torres, Stefan Lattner, Gael Richard, Singer Identity Representation Learning using Self-Supervised Techniques, Proceedings of the 24th International Society for Music Information Retrieval Conference (ISMIR 2023), https://doi.org/10.5281/zenodo.10265323

[11] Xulong Zhang, Jiale Qian, Yi Yu, Yifu Sun, Wei Li, Singer Identification Using Deep Timbre Feature Learning with KNN-Net, ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2021, http://dx.doi.org/10.1109/ICASSP39728.2021.9413774

[12] Pham Van Toan, Ngoc Ngo Quang Tran, Ta Minh Thanh, Deep Learning Approach for Singer Voice Classification of Vietnamese Popular Music, The 10th International Symposium on Information and Communication Technology (SoICT 2019), http://dx.doi.org/10.1145/3368926.336970

Received August 11,2025