LLB/TH/43/2025
TH6009

INTEGRATING MUSIC INFORMATION RETRIEVAL AND TRANSFER LEARNING FOR ADVANCED EMOTION RECOGNITION IN SRI LANKAN CROWD SOUNDSCAPES: DATASET CREATION AND ANALYSIS

P.B.S.N.Ariyathilake

219311M

MSc in Computer Science

Department of Computer Science & Engineering
Faculty of Engineering

University of Moratuwa
Sri Lanka

June 2025

INTEGRATING MUSIC INFORMATION RETRIEVAL AND TRANSFER LEARNING FOR ADVANCED EMOTION RECOGNITION IN SRI LANKAN CROWD SOUNDSCAPES: DATASET CREATION AND ANALYSIS

P.B.S.N.Ariyathilake

219311M

Thesis/Dissertation submitted in partial fulfillment of the requirements for the degree MSc in Computer Science

Department of Computer Science & Engineering
Faculty of Engineering

University of Moratuwa
Sri Lanka

June 2025

DECLARATION

I declare that this is my work and this thesis/dissertation does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other University or Institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature: *UOM Verified Signature* Date: 18.06.2025

The above candidate has carried out research for the PhD/MPhil/Master's thesis/dissertation under my supervision. I confirm that the declaration made above by the student is true and correct.

Name of Supervisor: Dr. Charith Chithraranjan

Signature of the Supervisor: Date: 18.06.2025

ACKNOWLEDGEMENT

First and foremost, I am deeply grateful to the University of Moratuwa and the Department of Computer Science & Engineering for providing me with the opportunity and resources to pursue my Master's studies.

I am deeply grateful to my supervisor, Dr. Charith Chitraranjan, for his continuous encouragement, guidance, understanding, and support throughout my research journey. His invaluable advice, constructive feedback, and inspiration greatly helped me to grow both theoretically and practically in this subject area. I am forever grateful for his mentorship.

I also extend my heartfelt thanks to Dr. Chathuranga Hettiarachchi, the MSc Coordinator, for his insightful discussions, advice, and support, which played an important role in the successful completion of this research.

A special word of appreciation goes to all the lecturers who taught during the MSc program. I am truly thankful for their dedication, time, and the invaluable knowledge they imparted.

I am also grateful to my friend Tharika and Shyamalee, whose willingness to share their own experiences helped me stay motivated and focused throughout this thesis journey.

I am deeply indebted to my parents, especially my mother, for constantly encouraging me to complete this research and for the unwavering support she has given, and to my sister for always checking on my progress, which motivated me to stay focused.

Finally, I would like to extend a heartfelt thank you to my husband for being a source of motivation and helping me in every means, where he can. And to my baby boy for the sacrifices made in sharing time with me during this journey. Their love and patience were my strength throughout.

ABSTRACT

Understanding crowd emotions through sound is critical for applications in event monitoring, public safety, and mental health studies. However, there has been a notable gap in the availability of specialized datasets and novel robust models for classifying crowd sound emotions. To address this, a comprehensive Sri Lankan Crowd Sound Emotion Dataset (SLCSED) was developed, enriched with detailed annotations, to support future research. The study proposes a computational framework based on Music Information Retrieval (MIR) techniques combined with advanced machine learning algorithms to perform emotion classification in crowd. Feature extraction was performed using MIR methods, Wav2Vec 2.0 embeddings, and Emotion2Vec representation. PCA was applied as a dimensionality reduction technique. Various machine learning and transfer learning classifiers, including TabNet, LightGBM, Multi-Layer Perceptrons (MLP), wav2vec, and emotion2vec, were evaluated. Specific architectures were tuned for better accuracy, such as LightGBM with Gradient boosting and MLPs with hidden layers of (128, 64) units. Furthermore, emotion recognition models were developed using supervised learning methods, drawing inspiration from approaches tested on decision trees, random forests, XGBoost, and LightGBM in related studies. The results demonstrated highly promising outcomes, with the LightGBM classifier achieving up to 99.95% validation accuracy on the Emotional Crowd Sounds Data(ECSD) dataset and the MLP achieving 99.53% on the SLCSED dataset without dimensionality reduction. PCA was found to slightly reduce the performance in most cases. Additionally, the Emotion2Vec framework showed significant improvements after PCA application, reaching 99.99% accuracy. These findings highlight the effectiveness of MIR-based feature engineering combined with carefully selected classifiers for crowd emotion detection. This work not only fills a major gap by introducing a localized and richly annotated dataset but also presents a robust methodological pipeline for crowd sound emotion recognition, paving the way for future applications in real-world monitoring and psychological analysis.

Keywords: ECSD, MIR, MLP**,** PCA, SLCSED

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Description
CNN	Convolutional Neural Networks
CSV	Comma Separated Values
CTC	Connectionist Temporal Classification
ECSD	Emotional Crowd Sound Data
FFT	Fast Fourier Transform
FN	False Negative
FP	False Positive
FP	Fourier parameter
GBDT	Gradient Boosting Decision Tree
GBM	Gradient Boosting Machine
KNN	K-Nearest Neighbors
LTD	Linear Discriminant Analysis
MFCC	Mel-frequency cepstral coefficients
MIR	Music Information Retrieval
MLi	Multi-Lingual
MLP	Multi-Layer Perceptrons
MTL	Multi-Task Learning
PANN	Pre-trained Audio Neural Network
PCA	Principal Component Analysis
PNN	Probabilistic Neural Network
RF	Random Forest
ROC	Receiver-operating characteristic
SER	Speech Emotion Recognition
SLCSED	Sri Lankan Crowd Sound Emotion Dataset
SVM	Support Vector Machine
TIM-NET	Temporal-aware bI-direction Multi-scale Network

TN	True Negative
TP	True Positive
ViT	Vision Transformer