EVALUATION AND ANALYSIS OF POOLING METHODS FOR AUDIO APPLICATIONS

E. M. S. J. Bandara

248088D

Master of Science: Major Component of Research

Department of Computer Science and Engineering
Faculty of Engineering

University of Moratuwa
Sri Lanka

November 2025

EVALUATION AND ANALYSIS OF POOLING METHODS FOR AUDIO APPLICATIONS

E. M. S. J. Bandara

248088D

Thesis submitted in partial fulfilment of the requirements for the degree
Master of Science: Major Component of Research

Department of Computer Science and Engineering
Faculty of Engineering

University of Moratuwa
Sri Lanka

November 2025

DECLARATION

I declare that this is my own work and this Thesis does not incorporate without acknowledgment 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 acknowledgment 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: Date: 12/01/2025

The supervisor should certify the Thesis with the following declaration.

The above candidate has carried out research for the Master of Science: Major Component of Research Thesis under my supervision. I confirm that the declaration made above by the student is true and correct.

Name of Supervisor:

Signature of the Supervisor: Date: 14 Jan 2026

ACKNOWLEDGEMENT

I would like to extend my heartfelt appreciation to my principal supervisor, Dr. T. Uthayasanker, for his invaluable guidance, steadfast mentorship, and unwavering support throughout the course of this research. His encouragement and insight were instrumental in navigating the challenges of this work, and his commitment to excellence continuously pushed me to elevate the quality of my efforts.

I am also deeply grateful to the esteemed members of the progress review committee, Dr. Thanuja Ambegoda and Dr. Charith Chitraranjan, for their thoughtful feedback and constructive advice. Their expertise and perspectives played a pivotal role in shaping the direction and refinement of this study.

We extend our sincere thanks to the staff of the Department of Electrical Engineering, University of Moratuwa, for providing a conducive learning environment and the necessary resources to undertake this research. Their support and dedication greatly contributed to the successful completion of this work.

Furthermore, I would like to acknowledge and appreciate the authors, researchers, and professionals whose work has been referenced in this study. Their seminal contributions and extensive research have served as a strong foundation for this project and have significantly enriched its quality and depth.

Finally, I wish to sincerely thank my family and friends for their constant encouragement, understanding, and belief in my abilities. Their unwavering presence has been a source of strength and inspiration, allowing me to persevere and remain committed to my academic goals.

ABSTRACT

Pooling plays a critical role in spectrogram-based audio analysis by determining how time–frequency information is compressed and preserved within deep learning models. However, existing evaluations of pooling methods primarily emphasize downstream task performance, such as classification accuracy, rather than directly assessing their ability to preserve key spectrogram features. Furthermore, existing pooling methods often fail to account for the non-stationary and multi-scale nature of audio signals. This study addresses these gaps by developing unified frameworks for understanding, evaluating, and advancing pooling mechanisms in audio analysis applications. We first propose a standardized evaluation framework using 17 metrics across four domains to assess pooling performance comprehensively. We then examine how specific pooling strategies affect feature retention across 15 audio application types. By identifying each application's key spectrogram characteristics and grouping them accordingly, we evaluate 12 pooling methods and analyze which features each method preserves. This provides task-level insight into selecting the most effective pooling technique, enabling more granular evaluation, improved explainability, and more efficient audio analysis pipelines. Building upon these insights, the study introduces a learnable adaptive wavelet pooling framework that combines multiscale feature decomposition with dynamic subband weighting, enabling the network to flexibly adjust the preservation of local and global features. To comprehensively evaluate performance and interpretability, we consider three major audio domains; speech, music, and environmental sounds and further analyze specific categories within each domain such as content, speaker/source, semantics, environmental factors etc. These categories allow systematic examination of how adaptive pooling responds to different information requirements, highlighting which spectral and temporal features are most critical for each type of application. Experimental results across representative datasets show that the proposed method consistently outperforms conventional and adaptive pooling baselines, achieving higher accuracy across all domains.

Keywords: Audio analysis, Spectrograms, Dimension Reduction, Wavelet Pooling, Evaluation Metrics

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Description
AST	Audio Spectrogram Transformer
AWP	Adaptive Wavelet Pooling
CNNs	Convolutional Neural Networks
CQT	Constant-Q Transforms
CRNNs	Convolutional Recurrent Neural Networks
DFT	Discrete Fourier Transform
DWT	discrete wavelet transform
FFT	Fast Fourier Transform
LPC	Linear Predictive Coding
LSTM	Long Short-Term Memory
MFCCs	Mel-Frequency Cepstral Coefficients
RBMs	Restricted Boltzmann Machines
RNNs	Recurrent Neural Networks
SAWP	Scaled Adaptive Wavelet Pooling
STFT	Short-Time Fourier Transform
SWP	Scaled Wavelet Pooling
ZCR	Zero-Crossing Rate