Corrosion Management ISSN: 1355-5243
(https://corrosion-management.com/)
Volume 35, Issue 02 – 2025
SCOPUS

CORROSION DETECTION IN CONCRETE USING PHASE-BASED SPEECH ENHANCEMENT AND RECOGNITION TECHNIQUES WITH AI FRAMEWORK

Nigamananda Mishra[1], Tusar Kanti Dash[1], Jagannath Dayal Pradhan[1] and Ganapati Panda[1]

Abstract
Automatic Speech Recognition (ASR) is one of the oldest signal processing techniques which deals with detecting different types of speech including spoken words, emotions, disease, speaker identity, languages, music, and noise levels. Recently, these standard ASR methods have been adopted into the fields of acoustics, bioinformatics, wildlife, and environmental audio detection. One of the upcoming applications of ASR is in civil engineering which deals with the hammer sound test. This test is one of the non-destructive tests which is performed to detect the concrete quality and surface defects. This research problem has been taken in this paper and investigated with various phase-based audio and statistical features along with speech enhancement techniques to reduce unwanted noises. Phase of the audio signal plays a crucial role in extracting relevant audio information which is less explored in the audio signal processing. The proposed phase-based method has been tested with standard hammer sound datasets and other environmental sound datasets. It has been observed from the analysis of the simulated results that the proposed method has been demonstrating superior performance.

Keywords
Corrosion Detection, Concrete Structures, AI-ML, Speech Recognition, Audio features

Introduction

Since the last decade, speech recognition techniques have experienced considerable growth due to the evolution and application of artificial intelligence in audio signal processing[1] In continuation to the development of the traditional ASR models, there has also been extensive use of deep learning models. The operation of automatic speech recognition depends on large corporate datasets with extensive processing power and storage capacity[2]. The combination of audio along with visual signals have shown improvements in the recognition results. The extensive review of audio-visual speech recognition techniques built since 2010 demonstrates the success of these combination methods in developing more resistant and advanced systems[3]. ASR models have been used for various applications excluding speech processing like avian monitoriing, water signal processing, acoustic analysis, and robotics. In one application of ASR vocalized speech has been converted into written text. This has been applied mainly in adult speech patterns. Recently this study has been extended to the child sound patterns[4].

Recently, ASR has been used in civil engineering with several signal processing schemes. The assessment of material damage requires acoustic emission signal processing through parameter examination and waveform analysis[5]. The paper first introduces the importance of parameter analysis before moving onto waveform analysis which focuses on wavelet transform along with its spectral analysis approach. Laboratory testing reveals that the wavelet transform along with Fourier Transform methods succeed in identifying material damage indicators as well as damage onset indicators. The poor condition of infrastructure creates an increased need for stable structural health monitoring systems. The early warning capabilities of Piezoelectric sensors and acoustic emission have been analyzed in[6]. The advanced detection systems use beyond parametric methods which specifically aid prestressed concrete examination. Two types of tests have been conducted to validate the system that showed its capability to detect relevant data properly. In another research, the nondestructive inspections of civil infrastructure have been done by studying time-frequency analysis of acoustic emission to monitor crack expansion[7].

Higher sampling rates are required along with prolonged observation durations in acoustic emissions. The designed methodology contains a dual operation which starts with data collection and follows with time-frequency inspection to detect frequencies that appear during structural failure events.

Corrosion Management ISSN: 1355-5243
(https://corrosion-management.com/)
Volume 35, Issue 02 – 2025
SCOPUS

The effectiveness of the system is confirmed through concrete structure testing standards. Another application of audio signal analysis in civil engineering is the proper identification and monitoring of construction heavy equipment. Several traditional methods exist for equipment tracking including direct observation and active sensors or computer vision systems. In this paper an audio detection system has been presented which utilizes particular equipment sounds to detect activities by analyzing recorded and processed audio recordings[8].

[1]Electronics & Communication Engineering, C V Raman Global University, Bhubaneswar, India Emails: nm4479@gmail.com (Nigamananda Mishra), ganapati.panda@gmail.com (Ganapati Panda)

Corresponding author:
Tusar Kanti Dash, Electronics & Communication Engineering, C V Raman Global University, Bhubaneswar, India
Email: tusarkantidash@gmail.com

The speech enhancement methods play a crucial role in improving the performance of ASR models[9]. Optimal enhancement processes combine different methods of noise reduction through traditional algorithms together with machine learning techniques. Speech enhancement works by extracting features while selecting them and classifying them as a method to enhance speech quality. In another paper, the review of signal subspace speech enhancement and its impact on ASR robustness against noise is studied[10]. Subspace filtering decomposes noisy speech into signal and noise components using a low-rank speech model and noise correlation estimates. The different speech recognition and enhancement techniques are reviewed and analyzed because speech enhancement functions as an initial step which better enables ASR applications[11]. The natural signals recorded from microphones are usually contaminated by noise and reverberation which degrades recognition performance. The speech processing applications employ neural networks with Transformer-based models for spectral masking operations. Speech clarity improves for better recognition through the combination of spectral subtraction, Wiener filtering, and adaptive noise reduction methods.

The corrosion behavior of steel reinforcement within concrete structures represents a primary engineering issue during operations within challenging environments. Still, the resaerch in this domain has not yielded sufficient improvements to address the issue effectively[12]. Direct electrochemical and physical techniques and indirect damage-based evaluations represent the methods used for detection. The protection measures for concrete consist of two categories where prevention includes fiber-reinforced cementitious composite overlay alongside coatings and inhibitors while therapy includes cathodic protection together with electrochemical chloride extraction. In another paper, an investigation of multiple difficulties regarding steel corrosion in concrete which affects the economic framework and education sector has been studied[13]. Insufficient funding enables research and education efforts on corrosion even though it creates substantial financial effects. The main difficulties in maintenance of aging concrete structures involve economical sustainability combined with longlasting new construction design requirements. The corrosion testing mechanism is concrete is divided into two types including destructive and non-destructive. Non-destructive evaluation techniques are safer and less disruptive than destructive methods by reducing costs and testing time. Hammer sounding is a cost-effective method for detecting early corrosion defects in concrete structures[14].

Detection of concrete corrosion happens through the use of acoustic impulse-response and impact-echo methods without damaging the material. Both methods serve different purposes in concrete assessment where impact-echo works best in spot boundary detection. The impulse-response detects troublesome areas throughout large flooring areas[15]. In another study, hammering response analysis through sequential online machine learning which attains nearhuman competency in concrete structure evaluation. Realworld changes become possible for the proposed system which updates itself through sequential processes. The developed system operated through two sequential stages which included feature extraction followed by model updating since it processed 10,940 expert-annotated samples. The experimental data revealed high assessment excellence together with efficient operation and minimal computational needs[16].

It has been observed from the literature review that corrosion detection in concreter through speech recognition is an evolving and challenging research domain. In these ASR models, the speech enhancement schemes can be used for improving the accuracy of detection. These problem have been taken in this paper with the research objectives as listed below.

- Analysis of several audio features to be used in corroison detection along with the statistical feature extraction methods.
- Application of speech enhancement schemes along with ASR techniques to improve the detection accuracy
- Analysis and performance comparison of the proposed model with baseline models, and features

Corrosion Management ISSN: 1355-5243
(https://corrosion-management.com/)
Volume 35, Issue 02 – 2025
SCOPUS

The remainder of this paper is structured as follows. The specifications of the datasets and methodologies are covered in Section II. Section III presents the experimental setup and analysis of the experimental findings. Finally, Section IV discusses the conclusion and the future scope.

Materials and methods

Dataset

This dataset contains the hammering sounds of the concrete bridge downloaded from[17]. This dataset has two categories of sounds from the context of degradation in concrete bridges including Normal and Abnormal. A total of two hundred sound files are used with each category having approximately hundred sound files. The standard audio data augmentation techniques are applied to im;prove the training of AI models. Time stretching, pitch shifting, volume gain, and time shifting schemes are using under the audio augmentation[18]. By using the augmentation schemes, the number of audio samples are of five hundred in each category. The time domain representation of the signal and the corresponding spectrograms are represented in Figures 1 and 2.

Cepstral Features

The effectiveness of capturing signal spectral characteristics in speech and audio processing makes cepstral features a widely used method by professionals. Signal periodic components can be analyzed through cepstral evaluation which involves calculating Inverse Fourier transform from signal of spectral domain[19]. The speech recognition field together with speaker identification and audio classification often utilize these cepstral features that demonstrate superior

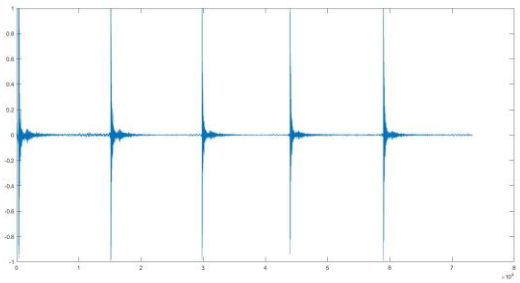

Figure 1. Time Domain Representation of the hammer sound

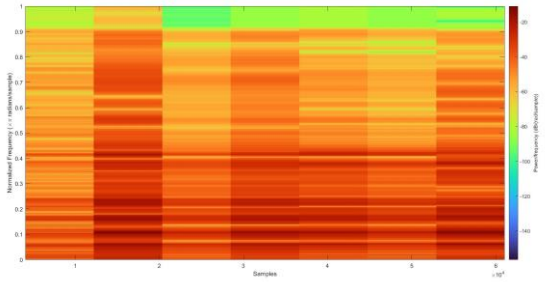

Figure 2. Spectrogram Representation of the hammer sound

efficiency. The process of extracting cepstral features involves several key steps:

- Signal Segmentation, Pre-processing, and Energy Calculation: The input signal is divided into smaller frames, pre-processed to enhance relevant features, and its energy is computed.
- Calculation of the Discrete Fourier Transform (DFT): The segmented signal undergoes a frequency-domain transformation using the DFT, allowing for spectral analysis.
- Conversion between Linear Scale and Perceptual Scale: The frequency components are mapped from a linear scale to a perceptually relevant scale, improving feature representation for human auditory perception.
- Application of Triangular Overlapping Filters: A series of triangular filters are applied to smoothen the spectrum and emphasize critical frequency bands.
- Application of Logarithm and Discrete Cosine Transform (DCT): The logarithm function is applied to capture amplitude variations, followed by the DCT to decorrelate features and obtain a compact representation of the spectral information.

This multi-step process ensures that cepstral features effectively capture the most relevant audio characteristics, making them highly valuable in various signal processing applications.

Statistical Features

In speech recognition, the statistical features play a crucial role as the sample levels are to be sorted from the frame level features In the current implementation the extracted statistical features are: mean, median, maximum, minimum, standard deviation, skewness, Kurtosis[20]. Trditionally, the speech signals are processed through various pre-processing techniques including framing and windowing. This is done to make the speech signal stationary which is quasi stationary in nature.

Corrosion Management ISSN: 1355-5243
(https://corrosion-management.com/)
Volume 35, Issue 02 – 2025
SCOPUS

In the proposed implementation, the phase-based cepstral features are extracted frame wise and the statistical features are extracted[21].

Improved phase-aware speech enhancement

Speech enhancement plays a crucial role in improving the performance of speech recognition algorithms. The modification of the phase in noisy speech signals plays a vital role in speech enhancement. However, in many speech enhancement algorithms, only the magnitude of the noisy speech is used and phase remains uncahnged. Recently, several speech denoising algorithms have been developed that leverage phase information, with the scaling factor being determined based on the estimated noise level. An effective bio-inspied speech ance ment algorithm has been developed by using the scaling factor and the accuracy of noise level estimation[22]. A neural network-based methodology for precisely estimating the scaling factor from the noise level is presented in this research. To find the ideal scaling factor for every noise level, the suggested method uses the firefly algorithm, a well-known bio-inspired optimization methodology. Furthermore, a precise correlation between the scaling factor and noise levels is established using an artificial neural network based on trigonometric functional expansion. An efficient method for predicting nonstationary noise is incorporated into the suggested algorithm to further improve the model's performance. The improved phase aware speech enhancement algorithm is used in the proposed method before extracting the features.

Experimental Results

Simulation analysis is conducted in this section to properly evaluate the effectiveness of the proposed model in corrosion detection in concrete. For all the classification tasks, stratified five-fold cross validationn scheme is used for training and testing. In this case, the dataset is split into five equal subsets at random, with four of the subsets being used for training and one for testing. To guarantee that every subgroup is used once for testing, this experiment is run five times. The evaluation measures are listed below:

$$\begin{aligned} Classification\ Accuracy &= \frac{TP + TN}{TP + TN + FP + FN} \\ Precision &= \frac{TP}{TP + FP} \\ Recall &= \frac{TP}{TP + FN} \\ F-1\ Score &= \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} \end{aligned} \quad (1)$$

Here, TP is the number of hammer samples that are predicted accurately as corrosive, TN is the number of non-corrosive samples that are being detected as non-corosive, FP is the number of non-corrosive samples detected as normal samples, and FN is the number of normal samples detected

Corrosion Management
ISSN: 1355-5243
(https://corrosion-management.com/)
Volume 35, Issue 02 – 2025
SCOPUS

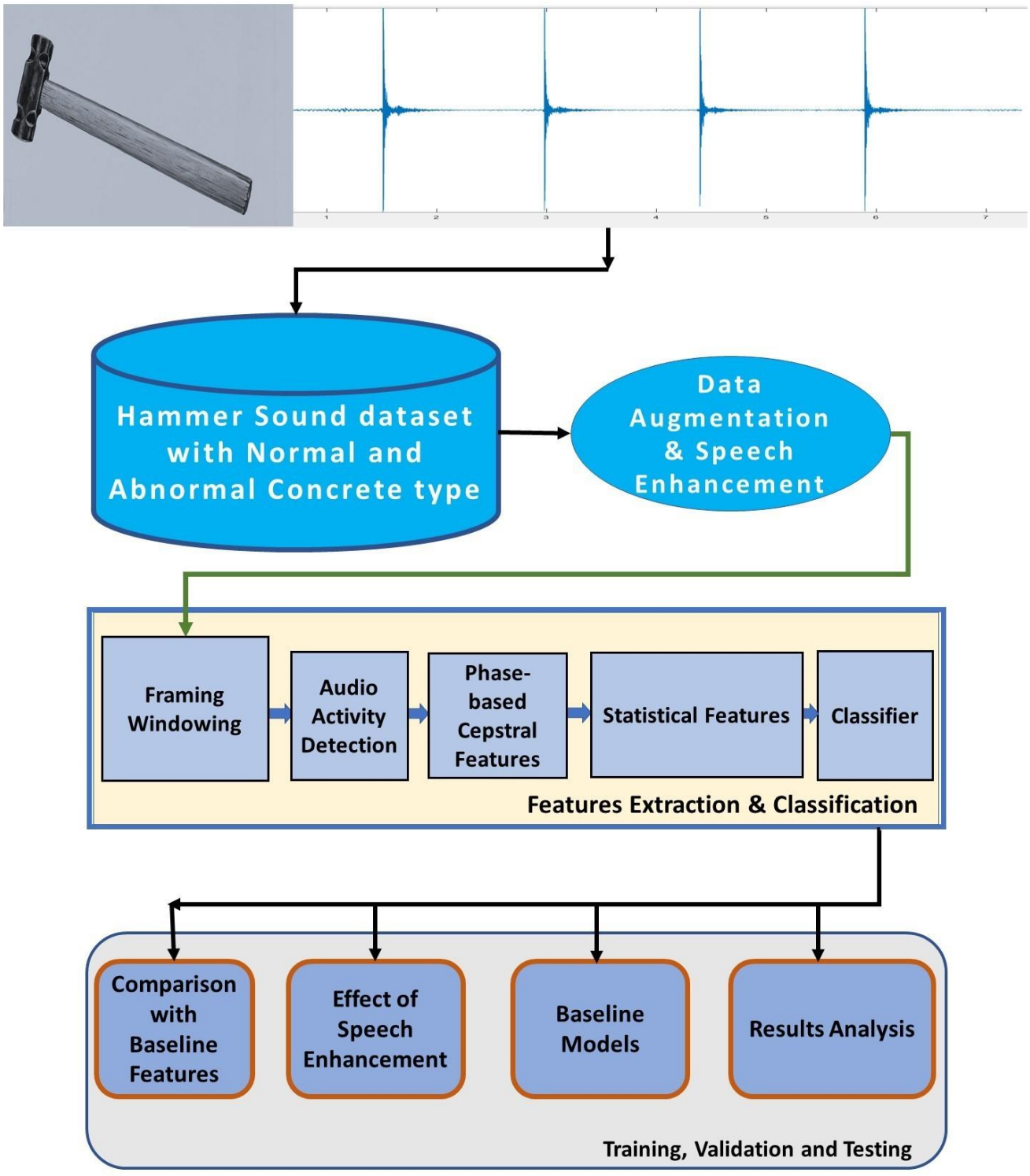

Figure 3. Block Diagram of the Implementation

as corrosive. The results are analysed from the context of baseline features, and models.

Corrosion Management ISSN: 1355-5243
(https://corrosion-management.com/)
Volume 35, Issue 02 – 2025
SCOPUS

Comparison with Standard Feature Extraction Techniques

At the first step, the proposed speech recognition scheme has been compared with standard feature extraction schemes including Mel-frequency cepstral coefficients (MFCC)[23] , Equivalent Rectangular Bandwidth (ERB) scale cepstral features[24] , and Bark scale features[25]. In the proposed implementation, the MFCC features with statistical features and phase aware speech enhancement scheme is used. Initially, the clean sound dataset has been added with ambient noise to make the noisy sound dataset. In all the models, the standard suppoty vector machine (SVM) classifier has been used. The results are being listed in Table 1.

Table 1. Perormance comparison with baseband features

Evaluation Mea-sures	MFCC	ERB	Bark	Proposed
Classification Accuracy	0.87	0.86	0.84	0.91
Precision	0.87	0.86	0.84	0.91
Recall	0.87	0.86	0.83	0.90
F-1 Score	0.86	0.86	0.83	0.90

It can be observed from the results analysis that the proposed model is performing comparatively better than other three feature extraction schemes. The MFCC deatures are perfoming at the second best position while ERB and Bark scale features are at the third and fourth position respectively. One of the probable reasons for the better perfromance of the proposed model is the additional use of two blocks including extended statistical features along with the speech enhancement schemes.

Comparative Analysis with Other Datasets

To further evaluate the effectiveness of the proposed model and its generalization capabilities, the model has been tested in another additional dataset called Environmental Sound dataset[26]. This dataset is derived from the Environmental Sound Classification dataset, which is widely used for environmental sound analysis and machine learning applications. It encompasses a diverse range of real-world sounds categorized into interior/domestic sounds and natural soundscapes & water-related sounds. The interior/domestic sound category includes common household noises such as door knocks, mouse clicks, keyboard typing, wooden door creaks, can openings, washing machines, vacuum cleaners, clock alarms, clock ticks, and glass breaking. These sounds are frequently encountered in indoor environments and are crucial for applications such as smart home automation, sound event detection, and acoustic scene analysis. The natural soundscapes & water-related sound category comprises environmental and atmospheric sounds such as rainfall, sea waves, crackling fire, crickets chirping, birds singing, water drops, wind blowing, pouring water, toilet flushes, and thunderstorms. These sounds are often studied for applications in ecological monitoring, environmental noise classification, and sound-based relaxation systems. In total, the dataset consists of 20 distinct sound categories, with each category containing 40 high-quality audio samples, each lasting 5 seconds. The dataset serves as a valuable resource for training and evaluating models in environmental sound classification, speech and audio processing, and various realworld applications requiring accurate sound recognition. Similar the the baseline feature analysis, the proposed model has been compared with MFCC, BARK, and ERB features with the speech enhancement scheme. It can be obseved that the proposed model has been working better the baseline features. The reason foir the superior performance is the use of statistical features in the proposed model while in the other cases, only few statistical parameters are included.

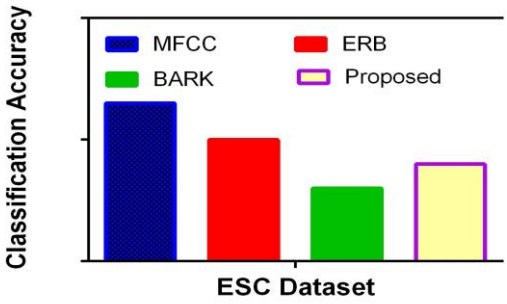

Figure 4. Comparative Analysis with ESC Dataset

Statistical Analysis

For the statistical analysis of the proposed model with the standard baseline models, the t-statistic values between two sets of classifiers are listed in Table 2[27]. The baseline models include the standard audio feature extraction techniques MFCC, ERB, and Bark along with the standard SVM classifier model. All these three models have been compared with the proposed feature extraction technique which includes the statistical additional features along with the standard SVM classifier. Each two classifier comparison results are analyzed separately. From the Table 2, it has been observed that most of the values are positive which indicates the superior performance of the proposed model in comparison to the baseline models.

Corrosion Management ISSN: 1355-5243
(https://corrosion-management.com/)
Volume 35, Issue 02 – 2025
SCOPUS

Table 2. Statistical comparison using t-static values

Proposed Model vs	Concrete dataset	ESC dataset
MFCC	1.2	0.8
ERB	1.7	1.6
Bark	0.9	1.3

Effect of Speech Enhancement

One of the important component of the proposed model is the use of additional sppech enhancement algorithm to enhance the effectiveness of the detection accuracy. As for the corrosion testing, it is not always possible to create the studiolike controlled envioronment, so there is a chance of the enviornmental noise affecting the overall performance. To evaluate the importance of the additional speech enhancement block, an ablation study has been performed with and without the use of speech enhancement (SE) on the corroison detection of the proposed and baseline models and shown in Figure 5. It can be observed that on an average 3% to 4% of performance improvement for the use of SE.

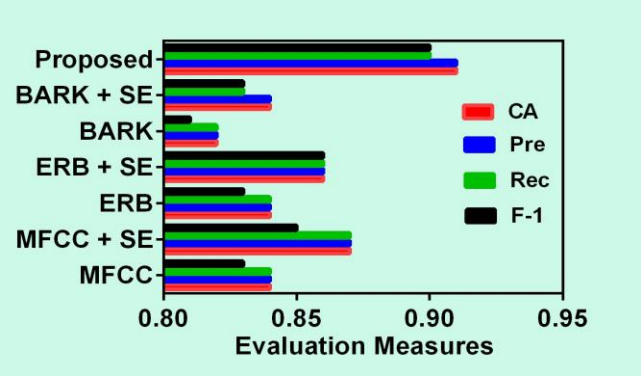

Figure 5. Effect of SE Algorithms

Conclusion

Detection of abnormality due to corrosion in concrete is crucial to enhance longevity and regular maintenance. For this purpose, two categories of detections are used including destructive and non-destructive. From the available several nondestructive methods, hammer sound detection is one of the popular and cost-effective methods. In this paper, this method has been implemented with standard speech recognition methods with statistical features and speech enhancement schemes. The method has been tested with two standard datasets and has been compared with baseline features. The proposed methods have demonstrated consistently superior performance than the baseline features mainly due to the extraction of relevant statistical features and speech enhancement schemes. In the future, the proposed detection scheme can be tested with multiple hammer sound datasets with real-time implementation.

Acknowledgements

The authors acknowledge 'National Seminar on Corrosion and its Prevention- Oil & Gas Industry' CPOG-2025, C.V. Raman Global University in collaboration with AMPP India Chapter.

References

1. O'Shaughnessy D. Trends and developments in automatic speech recognition research. *Computer Speech & Language* 2024; 83: 101538.
2. Kheddar H, Hemis M and Himeur Y. Automatic speech recognition using advanced deep learning approaches: A survey. *Information Fusion* 2024; : 102422.
3. Ivanko D, Ryumin D and Karpov A. A review of recent advances on deep learning methods for audio-visual speech recognition. *Mathematics* 2023; 11(12): 2665.
4. Bhardwaj V, Othman MTB, Kukreja V et al. Automatic speech recognition (asr) systems for children: A systematic literature review. *Applied Sciences* 2022; 12(9): 4419.
5. Zhao L, Kang L and Yao S. Research and application of acoustic emission signal processing technology. *Ieee Access* 2018; 7: 984–993.
6. Abdelrahman MA, ElBatanouny MK, Rose JR et al. Signal processing techniques for filtering acoustic emission data in prestressed concrete. *Research in Nondestructive Evaluation* 2019; 30(3): 127–148.
7. Lamonaca F and Carrozzini A. Monitoring of acoustic emissions in civil engineering structures by using time frequency representation. *Sensors & Transducers* 2010; 8: 42.
8. Cheng CF, Rashidi A, Davenport MA et al. Audio signal processing for activity recognition of construction heavy equipment. In *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*, volume 33. p. 1.
9. Das N, Chakraborty S, Chaki J et al. Fundamentals, present and future perspectives of speech enhancement. *International Journal of Speech Technology* 2021; 24(4): 883–901.
10. Hermus K, Wambacq P and Hamme HV. A review of signal subspace speech enhancement and its application to noise robust speech recognition. *EURASIP journal on advances in signal processing* 2006; 2007: 1–15.
11. Hepsiba D, Vinotha R and Anand LDV. Speech enhancement and recognition using deep learning algorithms: A review. *Computational Vision and Bio-Inspired Computing: Proceedings of ICCVBIC 2022* 2023; : 259–268.
12. Hu JY, Zhang SS, Chen E et al. A review on corrosion detection and protection of existing reinforced concrete (RC) structures.

Construction and Building Materials 2022; 325: 126718.

13. Angst UM. Challenges and opportunities in corrosion of steel in concrete. *Materials and Structures* 2018; 51(1): 4.
14. Zaki A, Chai HK, Aggelis DG et al. Non-destructive evaluation for corrosion monitoring in concrete: A review and capability of acoustic emission technique. *Sensors* 2015; 15(8): 19069–19101.
15. Hola J, Sadowski L and Schabowicz K. Nondestructive identification of delaminations in concrete floor toppings with acoustic methods. *Automation in Construction* 2011; 20(7): 799–807.
16. Ye J, Kobayashi T, Iwata M et al. Computerized hammer sounding interpretation for concrete assessment with online machine learning. *Sensors* 2018; 18(3): 833.
17. Emoto H, Baba Y, Asano H et al. Comparison of AI method on hammering sounds at concrete bridge. *Journal of Intelligence, Informatics and Infrastructure* 2020; 1: 1.
18. Salamon J and Bello JP. Deep convolutional neural networks and data augmentation for environmental sound classification. *IEEE Signal processing letters* 2017; 24(3): 279–283.
19. Rabiner L and Schafer R. *Theory and applications of digital speech processing*. Prentice Hall Press, 2010.
20. Dash TK, Chakraborty C, Mahapatra S et al. Gradient Boosting Machine and Efficient Combination of Features for Speech-Based Detection of COVID-19. *IEEE-JBHI (IEEE Transactions on Information Technology in Biomedicine)* 2022; .
21. Chakraborty C, Dash TK, Panda G et al. Phase-based Cepstral features for Automatic Speech Emotion Recognition of Low Resource Indian languages. *Transactions on Asian and LowResource Language Information Processing* 2022; .
22. Dash TK, Solanki SS and Panda G. Improved phase aware speech enhancement using bio-inspired and ANN techniques. *Analog Integrated Circuits and Signal Processing* 2020; 102(3): 465–477.
23. Prabakaran D and Sriuppili S. Speech processing: MFCC based feature extraction techniques-an investigation. In *Journal of Physics: Conference Series*, volume 1717. p. 012009.
24. Radha K and Bansal M. Towards modeling raw speech in gender identification of children using sincNet over ERB scale. *International Journal of Speech Technology* 2023; 26(3): 651–663.
25. Boualoulou N, Drissi TB and Nsiri B. Comparison of feature extraction methods between MFCC, BFCC, and GFCC with SVM Classifier for Parkinson's Disease diagnosis. In *International Conference on IoT Based Control Networks and Intelligent Systems*. pp. 231–247.
26. Piczak KJ. ESC: Dataset for environmental sound classification. In *Proceedings of the 23rd ACM international conference on Multimedia*. pp. 1015–1018.
27. Wong TT. Parametric methods for comparing the performance of two classification algorithms evaluated by k-fold cross validation on multiple data sets. *Pattern Recognition* 2017; 65: 97–107.