International Journal of Scientific and Technical Advancements

ISSN: 2454-1532

Application of Machine Learning in Dhwani Vigyan for Estimating Weather Conditions

Arfana Chowdhary[1,#], Parveen Kumar Lehana[2]

[1,2]Department of Electronics, University of Jammu, Jammu, J&K, India-180006

[1,#]arfanaliaquat@gmail.com

Abstract**—** In India, the study of sound, known as Dhwani Vigyan, has deep historical roots, with knowledge being passed down through generations. In ancient times, experts were adept at identifying the location of animals, objects, or even enemies simply by listening to the sounds they made. Arjuna, for instance, demonstrated extraordinary skill in hitting a target without sight, relying entirely on sound—this technique is called Shabda Bhedi Vidya. Ancient Indian literature contains numerous hymns, which were believed to have significant effects on both humans and the environment. Some hymns were thought to bring about positive outcomes, while others were considered to have negative consequences. In general, it is understood that distracting or unwanted sounds, often referred to as noise, can have harmful effects and are considered devoid of meaningful information. This paper explores the possibility of extracting meaningful information from such sounds, even when they seem chaotic or disruptive, using deep learning techniques. Take, for example, the sound produced by vehicles like bikes, cars, or buses on various types of roads. The noise varies depending on factors such as the speed of the vehicle, its type, and the condition of the road—whether it is dry, wet, or uneven. While such sounds are typically regarded as noise, they may actually carry valuable information. By analyzing the sounds of vehicles, it might be possible to predict environmental conditions, such as the type of road or weather effects on the road. This study aims to determine whether seemingly disruptive or noisy sounds could hold useful information, potentially revealing insights into the surrounding environment.

Keywords**—** Dhwani vigyan, shabda bhedi vidya, sound identification, sensory awareness, environmental conditions, noise analysis, deep learning.

I. Introduction

Dhwani vigyan is one of the most important field in acoustics. Throughout history, there have been individuals with remarkable skills, capable of interpreting sound to make precise judgments about objects, animals, and environments. The concept of sound as a powerful and transformative force is deeply rooted in various tales, illustrating its impact on perception, skill, and spirituality. One such story is that of King Pandu, who, while hunting, tragically misjudged an auditory cue. Mistaking the sounds of animals for their true nature, he accidentally killed Sage Kindama and his wife, when they were mating imitating the dear behaviour during sexual intercourse. This fatal error, driven by incomplete sensory perception, led to Pandu being cursed that “he would die if he ever touched his wife with sexual desire”. This tale underscores the dangers of relying solely on sound for decision-making, highlighting how incomplete or deceptive auditory inputs can lead to grave consequences [1].

In contrast, the mastery of Shabda Bhedi Vidya (the skill of targeting based on sound) by Arjuna, under the tutelage of Guru Dronacharya, demonstrates the immense potential of sound vigyan when honed through discipline and focus. Arjuna's training involved targeting a bird based entirely on sound while blindfolded, a feat that surpassed the abilities of his peers. His skill was later reaffirmed during the Pandavas' exile, when he successfully struck a wild boar guided only by its sound. Unknown to him, Lord Shiva, disguised as a hunter, had also targeted the same boar. This encounter led to Arjuna earning the divine Pashupatastra, a powerful celestial weapon.

Arjuna’s accomplishments symbolize the transformative power of sound as a tool for precision and mastery. His journey illustrates how dedication and training can elevate sound into a profound medium of skill, leaving a lasting legacy in the art of archery and beyond [2].

Similarly, the story of Shravan Kumar illustrated how sound could guide and sustain devotion but also lead to tragedy. Shravan story demosntrated heightened auditory senses to navigate safely when he was taking his blind parents for pilgrimage. However, his life ended tragically when King Dasharatha mistook the sound of water disturbed by Shravan and fatally shot him using Shabda Bhedi Vidya. Despite his suffering, Shravan forgave the king, asking him to care for his parents. This tale underscored sound’s dual nature as a tool for survival and a source of misfortune, demonstrating the ethical responsibility tied to its use. [3]. Tansen, the legendary musician in Akbar’s court, demonstrated his mastery over Naad Vigyan (sound science) through his extraordinary music. His ragas were believed to influence nature itself—he reportedly sang Raga Deepak to light lamps and Raga Megh Malhar to summon rain. These tales illustrated the mystical connection people once believed existed between sound, music, and the natural world, showcasing the profound power attributed to his art [4]. Orpheus, the musician and poet of Greek mythology, was said to have charmed humans, animals, and even inanimate objects with his music. When his wife Eurydice died, Orpheus used his melodious lyre to persuade the gods of the underworld to release her. His story served as a testament to the emotional and transformative power of sound, reflecting its ability to evoke deep feelings and create change [5].

Lord Shiva’s damaru (a small drum) was believed to be the source of cosmic sound in Hindu literature. The vibrations from the damaru were said to have created all the sounds and

Arfana Chowdhary and Parveen Kumar Lehana, “Application of Machine Learning in Dhwani Vigyan for Estimating Weather Conditions,” *International Journal of Scientific and Technical Advancements*, Volume 9, Issue 1, pp. 1-8, 2025.

set the rhythm of creation itself, symbolizing the connection between sound and the divine. The story linked sounds with the origins of language, music, and the universe [6]. Similarly, Lord Ganesha was closely associated with Omkara, the primordial sound "Om." This vibration was considered the essence of creation and wisdom. Stories of Ganesha often emphasized how sound could transform, helping to attain knowledge and overcome obstacles. These narratives across cultures illustrated how sound—whether as music, language, or vibration—was revered for its ability to connect, transform, and evoke the deepest human emotions, reflecting diverse aspects of Sound Vigyan in the past [7].

Latest research also attributes enough importance to sound vigyan. For example, in [8], advanced signal processing techniques such as Adaptive Noise Cancellation, Blind Signal Separation (BSS), and Back Propagation Neural Networks have been explored to separate and identify sound sources in aircraft Cockpit Voice Recorders (CVRs). Experimental results demonstrated successful separation of mixed sound signals from an ATR-72 cabin, facilitating more accurate accident investigations by isolating critical sounds and identifying potential causes.

Additionally, a related study examined the impact of aircraft cockpit interior sounds, with and without Active Noise Reduction (ANR) headsets, on pilots' affective reactions and expected performance. The findings revealed that ANR improved performance expectations, with affective reactions to noise closely linked to these expectations. However, speech characteristics and flight phases (climb, descent, level flight) influenced ANR effectiveness. Low-activation sounds positively affected performance expectations, suggesting that noise characteristics significantly impact pilot performance. Both studies highlight the need for further research on noise mitigation and its effects in operational aviation environments to enhance safety and maintenance [9].

Dhwani Vigyan is a thriving field of research with applications spanning various domains such as engineering, medicine, meteorology, agriculture, and more. Researchers are exploring innovative ways to utilize sound in solving complex problems and advancing technology. In engineering, sound waves are used for non-destructive testing to detect structural flaws in materials and machinery. In the medical field, ultrasound technology is widely applied for diagnostic imaging, therapy, and even breaking down kidney stones. Additionally, sound-based techniques are being developed for advanced hearing aids and cochlear implants to improve the quality of life for individuals with hearing impairments.

In meteorology, sound waves are employed to study atmospheric conditions. Technologies like infrasound monitoring help detect severe weather events and natural disasters like tornadoes and volcanic eruptions. Agriculture also benefits from sound science, where acoustic sensors monitor soil conditions, pest activity, and crop health. Specific problems like the cocktail party problem, where separating individual sounds in a noisy environment is a challenge, are being addressed using advanced signal processing and artificial intelligence. In urban areas, traffic noise analysis aids in developing better city planning and noise reduction strategies. Seismology uses sound waves to study earthquakes, with acoustic sensors providing valuable data on seismic activities and helping in early warning systems. Sound science continues to evolve, opening new possibilities in various fields. Its integration with modern technologies like artificial intelligence, machine learning, and robotics is further enhancing its potential to address real-world challenges.

In this paper, machine learning has been used for predicting wet and dry weather conditions by analyzing traffic sounds. The investigations will assist drivers in managing speed and adapting to weather conditions effectively. This approach aims to enhance driver performance. To support this goal, we discuss current challenges and outline possible future directions throughout the paper. Two critical components datasets and evaluation methods are highlighted as essential for training and assessing the performance of these systems. A curated list of resources, including datasets, objective measures, and various approaches, is available at the following link [10]. This resource is intended to provide both beginners and experts with easy access to a comprehensive collection of relevant materials, fostering continued progress in this field.

The remainder of this paper is organized as follows: Section I provides an introduction. Section II offers a brief overview of the literature review. Section III discusses the materials and methods in detail. Section IV presents and analyzes the results of the study. Finally, Section V concludes the paper, summarizing the key concepts and suggesting potential directions for future research.

.

II. Literature Survey

Speech enhancement and separation are vital for improving speech intelligibility and quality in noisy environments, where traditional signal-processing methods often falter under complex conditions. Deep learning has emerged as a powerful approach, leveraging multimodal data by combining acoustic features (e.g., spectrograms, phase information) with visual cues (e.g., lip movements, facial expressions). Fusion techniques like simple concatenation and attention mechanisms enable effective integration of these modalities, and models are trained using loss functions such as mean squared error (MSE) with targets like clean spectrogram masks. Audio-visual systems consistently outperform audio-only counterparts, particularly at low signal-to-noise ratios (SNRs), with attention mechanisms addressing challenges like modality dominance. Evaluation metrics such as PESQ, STOI, and SI-SDR are standard benchmarks, yet progress is hindered by the lack of standardized evaluation protocols, computational limitations for real-time applications, and the scarcity of large-scale datasets. Future research focuses on end-to-end multimodal systems, robust models for diverse real-world conditions, and innovative fusion strategies to enhance applications like teleconferencing and hearing aids [11].

Arfana Chowdhary and Parveen Kumar Lehana, "Application of Machine Learning in Dhwani Vigyan for Estimating Weather Conditions," *International Journal of Scientific and Technical Advancements*, Volume 9, Issue 1, pp. 1-8, 2025.

This study tackles the challenge of enhancing speech intelligibility in non-stationary noisy environments without prior knowledge of noise or speech statistics, where traditional methods relying on ideal masks or extensive computations fall short for real-time applications. The proposed Blind Acoustic Mask (BAM) is a novel time-domain technique that adaptively detects and suppresses noise components using a robust standard deviation estimator, preserving speech intelligibility by selectively processing samples with lower noise proportions. BAM operates without prior noise statistics, significantly reduces computational complexity—requiring just 10% of the time of traditional binary mask methods—and maintains high speech quality. Evaluated against Ideal Binary Mask (IBM), Target Binary Mask (TBM), and Non-stationary Noise Estimation for Speech Enhancement (NNESE), BAM demonstrated comparable intelligibility gains to ideal masks with an average STOI improvement of 0.17 and outperformed TBM at higher SNR levels. Findings highlight BAM's effectiveness in improving speech intelligibility in dynamic noise conditions, offering competitive performance with lower computational costs and suitability for real-time applications. This approach provides a practical, low-complexity alternative to ideal masks and lays the groundwork for efficient speech enhancement techniques in applications like hearing aids, telecommunication systems, and voice-activated devices [12]

Continuous speech separation (CSS) for meeting-style audio streams poses challenges such as longer durations, variable speaker numbers, and the need for low-latency real-time processing, which traditional methods like DPRNN struggle to address due to high computational demands. To overcome these issues, this study introduces Skipping Memory (SkiM), a novel time-domain speech separation model designed for efficient long-sequence modeling and low-latency operation. SkiM features a skipping memory mechanism that effectively processes long feature sequences generated by a time-domain encoder with a small stride, reducing computational costs by 75% compared to DPRNN while maintaining high separation performance. The model achieves latency below 1 millisecond, enabling real-time applications, and extends LSTM capabilities for long-sequence tasks. In evaluations, SkiM demonstrated a 17.1 dB signal-to-distortion ratio (SDR) improvement in simulated meeting scenarios, outperforming DPRNN in online CSS tasks while operating efficiently on low-power devices. Its combination of high performance, low latency, and computational efficiency positions SkiM as a practical solution for real-world applications like teleconferencing and meeting audio preprocessing. By setting a benchmark for efficient real-time CSS systems, SkiM paves the way for advancements in long-sequence modeling and resource-constrained deployments, opening opportunities for telecommunication and audio enhancement technologies [13]
Background music noise significantly impacts Automatic Speech Recognition (ASR) systems like Voice Search, leading to degraded Word Error Rates (WER), particularly in real-world scenarios. This study presents a model-based music-speech separation approach using Gaussian Mixture Models (GMMs) trained on background music preceding speech input. By leveraging the music prologue, the system mitigates music interference and achieves an 8% relative WER improvement at a Signal-to-Noise Ratio (SNR) of 10 dB. Optimal performance occurs when GMM models are trained on approximately 6 seconds of music prologue, as the structured and repetitive nature of popular music limits further gains beyond this point. Moderate-sized GMMs with 8 components effectively capture music dynamics, demonstrating the feasibility of training on short prologues—a common scenario in Voice Search applications. This approach is potentially extendable to other non-speech background noises, such as TV or radio sounds. Limitations include performance saturation with longer music prologues and a focus on moderate GMM sizes, prompting future exploration of speaker model adaptation techniques and generalization to other noise types. This work offers a practical solution for enhancing ASR performance in noisy environments, providing valuable insights into optimal prologue length and efficient model design for robust ASR systems [14]

Speech separation, the task of isolating a target voice from mixed sound signals, is crucial for effective human communication but remains challenging due to the high dimensionality of inputs like the magnitude spectrum of mixed speech. This study proposes a model that combines Convolutional Neural Networks (CNNs) and an attention mechanism to enhance separation accuracy by leveraging their strengths in feature extraction and sequence preservation. CNNs extract low-dimensional spatiotemporal features from high-dimensional inputs, while the attention mechanism minimizes the loss of sequence information, improving temporal modeling. Compared to the baseline model DRNN-2 + discrim, the proposed model achieves notable performance gains, including a 0.27 dB improvement in Global Normalized Signal-to-Distortion Ratio (GNSDR) and a 0.51 dB improvement in Global Signal-to-Interference Ratio (GSIR). Joint optimization of CNNs and attention mechanisms allows for more effective spatiotemporal feature extraction and sequence data preservation, outperforming traditional RNN-based models. Simulation experiments validate its superior performance, with future directions focused on further optimization, testing across diverse datasets, and exploring advanced mechanisms like transformers for enhanced robustness. This approach underscores the potential of integrating CNNs and attention mechanisms to advance speech separation technologies, providing a foundation for more accurate and efficient solutions [15]

The rapid growth of IoT has led to an increase in noisy speech data from sensors, making single-channel speech separation in multi-speaker scenarios a significant challenge, particularly in unsupervised environments. This study introduces CNMF+JADE, a hybrid algorithm combining Convolutional Non-Negative Matrix Factorization (CNMF) with Joint Approximative Diagonalization of Eigenmatrix (JADE), alongside an adaptive wavelet transform for speech enhancement. CNMF+JADE effectively separates target speech from mixed signals with minimal training data, while the adaptive wavelet transform refines weak and distorted signals, improving quality and intelligibility. Tested on the TIMIT dataset, the method demonstrated robust performance in noisy, single-channel scenarios, outperforming traditional approaches and showing high adaptability. Compared to conventional methods, which struggle with nonlinear spatial and temporal structures, and deep learning models that require extensive data, CNMF+JADE offers a computationally efficient solution for unsupervised settings. The study also highlights the potential of integrating deep learning with CNMF+JADE to further enhance performance in complex environments. Future directions include exploring real-world IoT applications, improving robustness, and leveraging advanced deep learning techniques. This work provides a robust and generalizable solution for processing noisy speech data, addressing a critical need in IoT-enabled speech sensor systems [16].

The *Dhvanyāloka* by Ānandavardhana, along with Abhinavagupta's commentary *Locana*, revolutionized Sanskrit literary criticism by introducing and elaborating on the concept of *dhvani* (suggestive meaning), which transcends primary (literal) and secondary (figurative) meanings. Drawing inspiration from Bhartrhari's *sphoṭa* theory, Ānandavardhana emphasized interpreting an utterance as a unified whole, shaped by context, where subtle linguistic and emotional cues evoke meanings beyond explicit connotations. Abhinavagupta expanded on this, particularly focusing on the role of *dhvani* in evoking emotions (*rasa*), such as the *śṛṅgāra rasa* in poetic narratives like Śiva and Pārvatī's love story. In this narrative, *dhvani* operates through *vibhāvas* (stimuli) and *anubhāvas* (responses), exemplified by Manmatha's arrow arousing Śiva's emotions, classified as Asaṁlakṣyakrama *dhvani*. These works established that explicit meanings serve as a foundation for grasping deeper, suggestive meanings, akin to using a torch to reveal hidden depths. By formalizing *dhvani* as central to poetic expression, these texts laid the groundwork for Indian poetics, profoundly influencing the understanding and appreciation of literary aesthetics [17].

Indian poetics encompasses a rich tradition of theories, including Rasa (Bharata), Alankara (Bhamaha), Guna-Dosha (Dandin), Riti (Vamana), Dhwani (Anandavardhana), Vakrokti (Kuntaka), and Auchitya (Khemendra), each contributing uniquely to literary analysis. Rasa emphasizes aesthetic effects and emotional responses, while Alankara focuses on figures of speech. Guna-Dosha evaluates literary qualities, Riti examines stylistic values, Dhwani explores the suggestive power of words, Vakrokti highlights oblique expression, and Auchitya stresses propriety in composition. Anandavardhana's theory of Dhwani unified literary criticism by introducing Abhidha (literal), Lakshana (figurative), and Vyanjana (suggested) meanings, with the latter considered the soul of poetry. His emphasis on the reader's role in experiencing rasa and interpreting Vyanjana highlights the universal, transpersonal nature of suggested meaning. Indian theories like Rasa and Dhwani find parallels with Western concepts such as Mimesis, Katharsis, Reader Response Theory, and Deconstruction, as explored in G. B. Mohan Thambi's *The Response to Poetry* (1968). These theories remain relevant, offering a framework for interpreting both Indian and Western literature, enriching global literary understanding, and revitalizing appreciation for aesthetic pleasure in a materialistic age. For scholars, integrating Indian poetics into modern critical practices opens avenues for blending cultural insights and enhancing methodologies for literary analysis worldwide [18].

III. Material and Method

The methodology employed in this study encompasses data preparation, simulation, machine learning model development, training, testing, and evaluation. A well-structured approach was followed to ensure the effective classification of emotional states in audio samples. The process begins with database preparation, where audio files are systematically labeled and organized for analysis. Spectrogram visualization and feature extraction facilitate data simulation to capture meaningful patterns. A Bi-Directional Long Short-Term Memory (Bi-LSTM) model is designed and trained using extracted features, with appropriate hyperparameter tuning and validation techniques. The trained model undergoes rigorous testing, and performance is assessed using various evaluation metrics. The following sections provide a detailed explanation of each step involved in this methodological framework.

Database Preparation

The dataset utilized in this study consists of 2,082 audio files categorized into two distinct emotional classes: dry and wet. Each audio file was systematically labeled and stored in a structured format using a Pandas DataFrame. The dataset was collected from various sources and preprocessed to ensure consistency. To facilitate data organization, directory structures were traversed using `os.walk`, and labels were extracted from filenames based on predefined delimiters. This ensured efficient access and retrieval of the audio samples for subsequent analysis.

Simulation

To enhance the robustness of the dataset, real-world conditions were simulated through spectrogram visualization and feature extraction. Spectrograms were generated using Short-Time Fourier Transform (STFT) to visualize the frequency content of audio signals. The transformation was computed using `librosa STFT`, and the amplitude was converted to decibels via `librosa amplitude_to_db`. Additionally, `librosa.display.specshow` was used for visual representation. This step allowed for a better understanding of the spectral characteristics of the audio data and contributed to feature extraction for model training.

Machine Learning Model Development

Feature extraction is crucial for preparing model inputs. Mel-Frequency Cepstral Coefficients (MFCCs) and Chroma Features were extracted to capture the temporal and frequency-domain characteristics of audio signals. The MFCC extraction involved computing 40-dimensional features using `librosa.feature.mfcc`, followed by mean computation across time frames to obtain a fixed-length feature vector. Chroma features were derived using `librosa.feature.chroma_stft` and normalized for uniformity. The processed feature vectors were then expanded in dimensions and normalized to ensure compatibility with the selected model architecture.

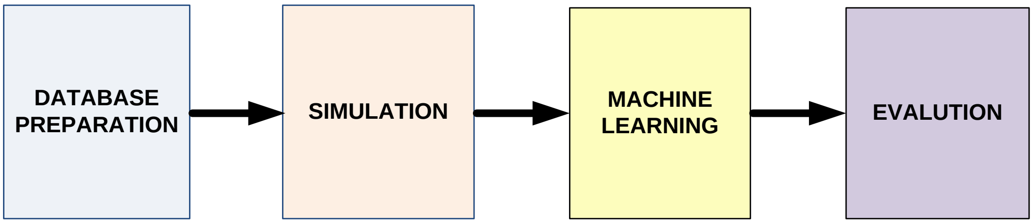

Fig.1 Block diagram.

Model Architecture and Training

A sequential deep learning model based on Bi-Directional Long Short-Term Memory (Bi-LSTM) was designed for sequence processing. The model architecture consisted of a Bi-LSTM layer with 256 units to capture sequential patterns, followed by fully connected dense layers with 128 and 64 neurons, respectively. Regularization techniques, such as dropout, were applied to prevent overfitting. The output layer utilized a softmax activation function for multi-class classification. The model was trained using categorical cross-entropy as the loss function and Adam as the optimizer. A validation split of 20% was used, and checkpointing mechanisms were implemented to save the best-performing model. A learning rate scheduler was also employed to optimize convergence.

Testing and Evaluation

The testing phase involved loading and preprocessing the dataset for inference. The dataset directory was recursively traversed to collect audio file paths and extract labels. A Pandas DataFrame was created to store this information, ensuring a structured approach to testing. The label distribution was visualized using Seaborn's histogram (`sns.displot`). To analyze audio samples, helper functions such as `waveplot` (for waveform visualization) and `spectrogram` (for spectral representation using STFT) were employed. Feature extraction was performed again to prepare the test samples, ensuring consistency with the training pipeline.

Model Inference and Performance Assessment

The preprocessed MFCC feature vectors were converted to NumPy arrays and reshaped for compatibility with the trained Bi-LSTM model. Labels were encoded using one-hot encoding. A pre-trained TensorFlow/Keras model was loaded for inference, and predictions were generated for the test dataset. The model's performance was evaluated using a confusion matrix visualized with a Seaborn heatmap. Additionally, accuracy and loss metrics were plotted across training epochs to assess the model's learning progress. The final trained model was saved in HDF5 format for future use, ensuring its reproducibility and applicability in real-world scenarios.

IV. Results and Discussion

The accuracy of estimating weather conditions from the given plot was analysed using the Bi-LSTM model, as described in the previous section. The shape of the plot is influenced by two primary conditions: wet and dry. The training and validation accuracy trends over 50 epochs reveal key insights into the model's learning behaviour. Initially, the training accuracy increases gradually, reaching a stable range of 80%-85%, indicating effective learning. The validation accuracy remains low in the initial epochs but experiences a sharp increase around epoch 15, suggesting the model has identified critical features in the dataset. Post this improvement, the validation accuracy stabilizes within 80%-90%, demonstrating strong learning capabilities.

The generalization gap, measured by the difference between training and validation accuracy, is minimal after epoch 15. This implies that the model generalizes well to unseen data without significant overfitting. The sudden jump in validation accuracy may be attributed to the optimizer or learning rate adjustments that enabled the model to recognize key patterns more effectively. The close alignment between training and validation accuracy in later epochs further suggests that the model is robust and capable of making accurate predictions.

Additionally, the training and validation loss curves provide further insights into the model's performance. The training loss steadily decreases, confirming that the model is effectively learning from the data. However, the validation loss exhibits fluctuations, with a noticeable spike around the 20th epoch, hinting at temporary instability or overfitting. Despite this, the validation loss stabilizes in the later epochs and trends closer to the training loss, suggesting improved learning stability as training progresses.

Arfana Chowdhary and Parveen Kumar Lehana, "Application of Machine Learning in Dhwani Vigyan for Estimating Weather Conditions," *International Journal of Scientific and Technical Advancements*, Volume 9, Issue 1, pp. 1-8, 2025.

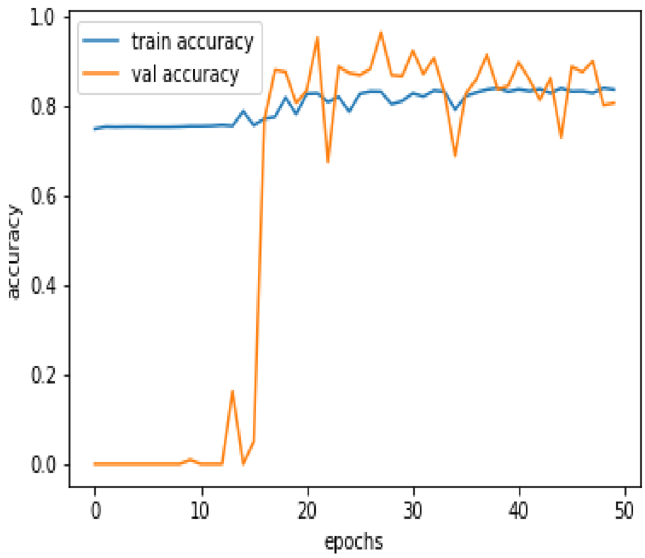

Fig.2. Effect of vehicle sound on the estimation training accuracy and validation accuracy

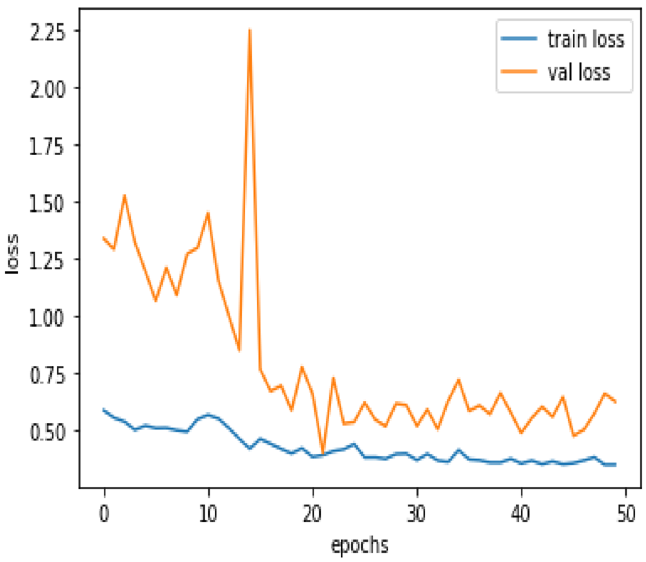

Fig.3. Effect of vehicle sound estimation of training loss and validation loss.

Overall, while the model demonstrates strong learning performance, the observed fluctuations in validation loss suggest potential areas for improvement. Employing additional regularization techniques, such as dropout or early stopping, could mitigate overfitting risks. Fine-tuning hyperparameters may also help enhance model stability and further optimize performance. These findings indicate that the Bi-LSTM model is a promising approach for weather condition estimation, with room for refinements to improve accuracy and consistency.

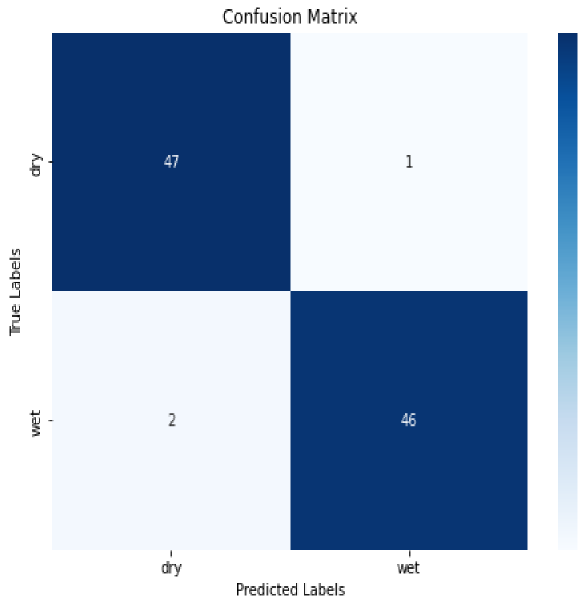

Fig.4. Confusion matrix for wet and dry prediction

The confusion matrix provides a comprehensive evaluation of the model's performance in classifying "dry" and "wet" labels. The model accurately classified 47 instances as "dry" (true positives) and 46 instances as "wet" (true negatives), demonstrating its effectiveness in distinguishing between the two conditions. However, it misclassified one "dry" instance as "wet" (false positive) and two "wet" instances as "dry" (false negatives). These minor misclassifications suggest that while the model performs well overall, further refinements in feature selection or training strategies may enhance its predictive accuracy.

Overall, the model demonstrates high accuracy and balanced performance across both classes, with minimal misclassification

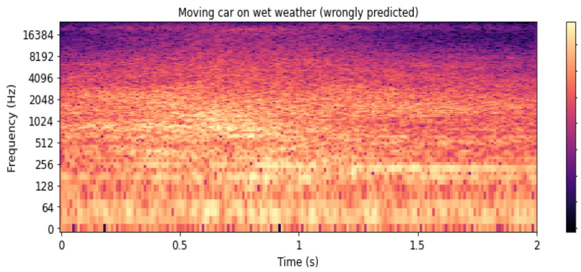

Fig.5spectrogram on wet weather condition wrongly predicted

Arfana Chowdhary and Parveen Kumar Lehana, “Application of Machine Learning in Dhwani Vigyan for Estimating Weather Conditions,” *International Journal of Scientific and Technical Advancements*, Volume 9, Issue 1, pp. 1-8, 2025.

The above image [fig.5] shows a spectrogram of an audio signal labeled "Moving car on wet weather (wrongly predicted)." The x-axis represents time in seconds (up to 2 seconds), while the y-axis represents frequency in hertz (Hz), ranging from 0 to 16384 Hz. The color intensity indicates the amplitude in decibels (dB), with warmer colors (yellow) representing higher amplitudes and cooler colors (purple) representing lower amplitudes, ranging from -70 dB to 0 dB. Wrongly prediction on wet weather means equally dense from both side and minute less from middle. The presence of high-frequency components and consistent patterns across time suggests that the model may have failed to correctly distinguish features of the "wet weather" sound, leading to a misclassification.

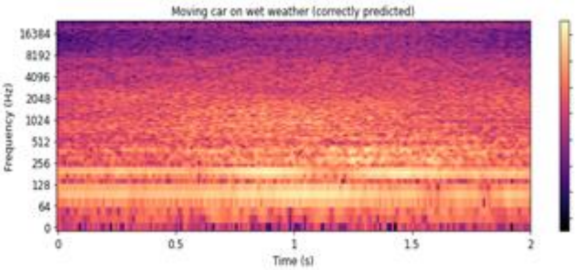

Fig.6 spectrogram on wet weather condition correctly predicted

The above spectrogram[fig.6] represents an audio signal labeled "Moving car on wet weather (correctly predicted)." it shows time (up to 2 seconds) on the x-axis and frequency (from 0 to 16384 Hz) on the y-axis. The color scale indicates amplitude in decibels (dB), with a range from -60 dB to +10 dB. Compared to the previous image, this spectrogram appears to have distinct, brighter bands around 128 Hz and 256 Hz, which might contain audio features that allowed the model to correctly classify it as "moving car on wet weather." Correctly prediction on wet weather means more dense on 0dB to -10dB.The spectral patterns in the higher frequency range remain consistent, suggesting similar environmental sounds or noise across both samples.

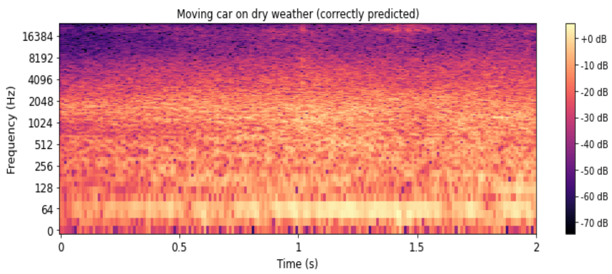

Fig.7spectrogram on dry weather condition correctly predicted

The spectrogram in fig.7 visualizes the frequency content over time for a correctly predicted audio clip of a moving car in dry weather conditions. The horizontal axis represents time in seconds, while the vertical axis shows frequency in Hertz (Hz). The color gradient indicates the intensity in decibels (dB), with brighter colors representing stronger frequencies. Correctly prediction on dry weather means more dense from 8192 Hz-16384 Hz. This pattern likely reveals characteristic audio features that helped the model accurately classify the sound.

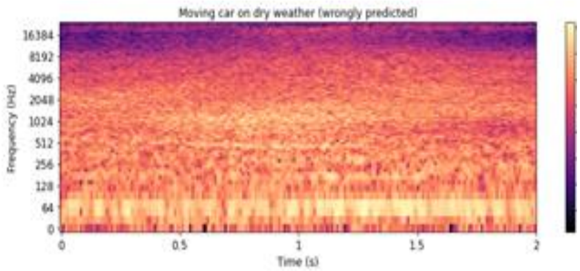

Fig.8. spectrogram on dry weather condition wrongly predicted

This spectrogram in fig.8 shows the audio signal labeled "Moving car on dry weather (wrongly predicted)." The x-axis represents time in seconds (up to 2 seconds), and the y-axis represents frequency in hertz (Hz), ranging from 0 to 16384 Hz. The color bar on the right indicates the amplitude in decibels (dB), ranging from -70 dB to 0 dB.In this spectrogram, the spectral pattern is similar to those of a moving car but lacks distinct low-frequency bands that might differentiate it from wet-weather sounds. Wrongly prediction on dry weather means less dense from 0dB to -10dB in middle of the spectrum. The energy distribution is relatively consistent across frequencies, which may have contributed to the incorrect prediction, as it resembles the "moving car on wet weather" pattern seen in the previous images.

V. Conclusion

The study successfully demonstrated the application of machine learning, particularly the Bi-LSTM model, in classifying weather conditions based on traffic sound analysis. The results confirmed a strong correlation between auditory signals and environmental conditions, reinforcing the effectiveness of sound-based classification in real-world scenarios. The Bi-LSTM model exhibited high accuracy with minimal generalization gaps, indicating its robustness in distinguishing between "dry" and "wet" conditions. However, occasional fluctuations in validation loss suggest opportunities for improvement through advanced regularization techniques and fine-tuning hyper parameters. This work contributes to the growing field of auditory-based machine learning, highlighting its potential for enhancing road safety, meteorological predictions, and environmental monitoring.

Future research can focus on expanding the dataset, incorporating additional audio features, and integrating real-time adaptive learning for enhanced performance. Exploring alternative deep learning architectures, such as transformers or CNN-LSTM hybrids, may further improve classification accuracy. Additionally, combining sound analysis with other environmental data sources, such as humidity or temperature sensors, could strengthen predictive capabilities. The methodology can also be extended to broader applications, including urban noise mapping and climate monitoring, to address larger-scale environmental challenges. By refining feature extraction techniques and optimizing model architectures, future studies can enhance the accuracy, efficiency, and applicability of auditory-based weather prediction models.

REFERENCES

[1] "Kindama," *Wikipedia*, https://en.wikipedia.org/wiki/Kindama.

[2] "Arjuna: The Legendary Archer in Indian Mythology," *Avid Archer*, https://avid-archer.com/arjuna-legendary-archer-indian-mythology/.

[3] "Shravana Kumara," *Wikipedia*, https://en.wikipedia.org/ wiki/Shravana_Kumara.

[4] "Tansen," *Britannica*, https://www.britannica.com/biography/Tansen.

[5] "The Story of Orpheus and Eurydice," *The Collector*, https://www.thecollector.com/orpheus-eurydice-story/.

[6] "Shiva's Damaru and Sanskrit," *Adhyatmic Views*, https://adhyatmicviews.wordpress.com/2018/03/02/shivas-damaru-and-sanskrit/.

[7] N. Krishna, "Sacred groves of India," *Seminar*, vol. 651, 2013. Available: https://www.india-seminar.com/2013/651/651_nanditha_krishna.htm.

[8] F.-B. Hsiao, S.-Y. Han, S.-C. Hsieh, and L. K. Wang, "Sound source separation and identification for aircraft cockpit voice recorder," *J. Aerosp. Comput. Inf. Commun.*, vol. 1, no. 12, pp. 466–483, Dec. 2004, doi: 10.2514/1.11266.

[9] J. Lindvall and D. Västfjäll, "The effect of interior aircraft noise on pilot performance," *Percept. Mot. Skills*, vol. 116, no. 2, pp. 472–490, Apr. 2013, doi: 10.2466/27.24.pms.116.2.472-490.

[10] "Dataset on environmental sound classification," *Zenodo*, https://zenodo.org/record/7551553.

[11] D. Michelsanti et al., "An overview of deep-learning-based audio-visual speech enhancement and separation," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 29, pp. 1368–1396, 2021, doi: 10.1109/TASLP.2021.3066303.

[12] F. Farias and R. Coelho, "Blind adaptive mask to improve intelligibility of non-stationary noisy speech," *IEEE Signal Process. Lett.*, vol. 28, pp. 1170–1174, 2021, doi: 10.1109/LSP.2021.3086405.

[13] C. Li, L. Yang, W. Wang, and Y. Qian, "SKIM: Skipping memory LSTM for low-latency real-time continuous speech separation," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, May 2022, pp. 681–685, doi: 10.1109/ICASSP43922.2022.9746372.

[14] T. Hughes and T. Kristjansson, "Music models for music-speech separation," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Mar. 2012, pp. 4917–4920, doi: 10.1109/ICASSP.2012.6289022.

[15] C.-M. Yuan, X.-M. Sun, and H. Zhao, "Speech separation using convolutional neural network and attention mechanism," *Discrete Dyn. Nat. Soc.*, vol. 2020, pp. 1–10, Jul. 2020, doi: 10.1155/2020/2196893.

[16] D. Jiang, Z. He, Y. Lin, Y. Chen, and L. Xu, "An improved unsupervised single-channel speech separation algorithm for processing speech sensor signals," *Wireless Commun. Mobile Comput.*, vol. 2021, no. 1, Jan. 2021, doi: 10.1155/2021/6655125.

[17] A. Singh, "Dhvani theory in Indian aesthetics," *Int. J. Eng. Lit. Soc. Sci.*, vol. 7, no. 3, pp. 233–237, 2022, doi: 10.22161/ijels.73.34.

[18] S. Sinha, "Unveiling the unspoken: A critical study of 'Dhwani' in Indian aesthetics," *J. South Asian Exch.*, vol. 1, no. 1, Aug. 2024, doi: 10.21659/JSAE/V1N2/V1N2022.