Deep Learning-Based Speech Enhancement for Robust Speech Recognition in Noisy Environments

Sowmya C S,
Department of Electronics and Communication Engineering ,
Presidency University, Bangalore, Karnataka, India,
sowmya.cs@presidencyuniversity.in

Neeraj Das,
Maharishi School of Engineering & Technology, Maharishi University of Information Technology, Uttar Pradesh, India,
dasium3000@gmail.com

Divya Sharma,
Centre of Research Impact and Outcome, Chitkara University, Rajpura- 140417, Punjab, India
divya.sharma.orp@chitkara.edu.in

Sanjoy Mondal,
Department of Computer Science and Information Technology, Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, Odisha, India,
sanjoymondal@soa.ac.in

Ishika Soni,
Chitkara Centre for Research and Development, Chitkara University, Baddi, Himachal Pradesh-174103 India
ishika.soni.orp@chitkara.edu.in

Naveen Kumar V
School of Mangement - PG Management JAIN (Deemed to be University), Bangalore, Karnataka
naveenkumar_v@cms.ac.in

Abstract — **It is a very important problem since voice-controlled devices and speech-to-text transcription are only two examples of how automatic recognition of spoken language in noisy environments may impact our lives. Still, speech recognition systems may only sometimes work effectively in background noise, which can result in errors and misinterpretations. This problem dealt with the deep learning model-based speech enhancement techniques for robust ASR in adverse environments. These methods employ deep learning to learn and deny the noisy speech signals, rendering it a better-suited input for recognition in a typical ASR system. Deep learning-based speech enhancement is a machine-learning system using a huge, loud, clean speech data dataset to train DNN. So, the network is trained to find a mapping from some noisy speech signals to ideally clean versions of these speech sounds by removing its background noise. This spectral representation of speech will now have reduced noise, which the ASR much more easily recognizes. These methods have significantly improved against traditional speech enhancement techniques in demising, eventually improving recognition performance. They can turn in high noise levels and are resilient against acoustics drifts. This makes them well-suited to real-world applications when noise levels can vary.**

Keywords— Mapping, Improving, Recognition, Significantly, Resilient

I. INTRODUCTION

From virtual assistants to automatic transcription, speech recognition has become integral to our lives[1]. That being said, in environments with noise, the performance dramatically decreases, resulting in errors and reduced accuracy. To mitigate this problem, the speech enhancement task was to pose noise speech signals closer to their clean version and improve ASR[2]. One very popular technique is deep learning-based Speech Enhancement. It is used for data where deep learning comes into play, and later, layers of artificial neural networks are used to learn features from complex and large amounts of data [3]. Since then, self-attention has become an increasingly popular technique due to the abundance of large-scale speech datasets and strong computational resources, leading it to be successful across different domains, including natural language processing (NLP) and computer vision[4]. In the speech enhancement field, deep learning can effectively enhance noisy signals to improve SR systems' robustness under noise conditions. Deep learning-based speech enhancement works because it can model the complex relationship between noisy and clean speech signals [5]. The conventional speech enhancement techniques, such as spectral subtraction, Wiener filtering, etc., are based on assumptions about the properties of noise and clean-speech signature. These assumptions can be, in a real-world scenario, assumptions do not hold, and your performance may be suboptimal [6]. In contrast, deep learning learns directly from the data with no assumptions and hence can cope better with a broad range of noises and speech signals. Noise reduction: One of the earliest use cases where deep learning started to be applied in speech enhancement[7]. This method involves training a deep neural network to learn the mapping of an activity classified as noisy speech sound with the same label being clean[8]. The network is optimized to deny the loud and clear signals by learning what features correspond well with speech [9]. The network produces a clean speech signal output, which can be used in a concatenated pipeline with any feature extractor like MFCC to get cleaner forms for language processing and recognition applications, resulting in better performance in noise environments[10]. Deep Learning for Speech Dereverberation: One other method is to eliminate the repercussions brought on by echoes triggered as an effect of this area environment utilizing heavy education. This can be helpful in use cases, such as telephone conferences or lectures[11]. However, due to the reverb, speech signals are often severely impacted in quality, and these systems need help deciphering these corrupted segments[12]. Dereverberation techniques typically employ deep learning-based methods for reverberant speech enhancement by estimating the room impulse response using convolution neural networks before and after removing the echoes from aged speech signals [13]. One of the more recent innovations for speech enhancement using deep learning is generative adversarial networks[14]. GANs are a pair of neural networks: the generator and discriminator of a deep learning model. The generator learns to produce convincing fake outputs while the discriminator tries to tell them apart from real inputs[15]. This creates competition between the two networks, improving the generated outputs' quality. GANs have been widely applied in speech enhancement for noise reduction and dereverberation tasks. Although deep learning-based speech enhancement has succeeded, it still encounters challenges. One key factor is the size and range of data that can be used for training[16]. While deep learning models need large amounts of data to train, obtaining and annotating this amount of data is costly in terms of time invested. This has resulted in methods that generate synthetic sounds for designing deep neural networks based on dereverberation approaches [17]. Abstract With the rise of deep learning, enhanced speech enhancement based on a single.', 'Sentence voice noise has become a powerful tool to improve accuracy for degraded conditions in noisy environments. This is great since MFCCs are modeled to capture steady speech signals, and my data contains a lot of smudgy noises from the recording [18]. Deep and complex architectures: With the sophistication of models increasing and better large dataset availability, deep learning-based SE will advance further, boosting accuracy/robustness for speech recognition performance in the real world. The main contribution of the paper has the following

- We, as such, provide two main contributions: We use deep learning techniques to enhance speech recognition accuracy in the presence of noise, and advanced neural network models are used to better separate speech from noise and improve speech transcription in difficult acoustic situations.
- Resistance to different kinds of noise: Another valuable input is that the model should also handle all types of background noises. We show for the first time that speech enhancement can also be meaningfully approached in deep learning, provided there is a diverse enough corpus of noisy environments to train on.
- The researchers are likewise working on a real-time processing system for speech signals. This is crucial for applications that require real-time speech recognition, video conferencing, or hands-free voice control since post-processing delays can harm user experience.
- Fully automated system: We provide a complete pipeline, an end-to-end solution requiring no further hand-crafted feature extraction or post-processing. This streamlines system architecture and decreases computational complexity, which in turn makes it more applicable to real-world problems.

II. Related Works

Noise Robust Speech Recognition- One of the most contentious problems within Artificial intelligence and Machine learning is recognizing speech in a noisy environment[19]. With the popularity of natural interfaces and voice-based systems/devices, so does the request for precise and strong speech recognition. In practice, the presence of noise drastically affects the performance of our speech recognition systems[20]. Deep learning-based Speech enhancement techniques are some proposed solutions for this problem. These techniques enhance speech signals, which will help reduce background noise and make a better and more accurate species recognizer [21]. However, this approach has many problems and limitations. A substantial difficulty in this area is the non-robust nature of deep learning-based speech enhancement algorithms[22]. The algorithms are usually trained and tested on datasets with clean speech, which does not reflect a true noise-generation process[23]. This means environments with many variable and unpredictable backgrounds may need more optimal performance the type and noise level can significantly impact how well these algorithms perform. The impact noise also performs differently[24]. Thus, it is important to keep the algorithms used for speech enhancement well-trained with many noise types. Another challenge is the high computational cost of deep learning speech enhancement. These algorithms need to be trained on huge forms of information, which can take masses of time throughout the computation trained models may be complex and require significant computational resources for real-time processing, making them inappropriate for deployment in low-compute capability applications [25]. While deep learning-based speech enhancement methods can extend non-stationary or non-Gaussian noise, these algorithms are expected to suffer from different degradation. Noise in real-world environments is typically non-stationary, which means it changes over time and often has a Gaussian/non-Gaussian distribution. This also makes it difficult for algorithms working with a type of noise to perform well in other noisy conditions. The more important one is that no generally accepted evaluation metrics for speech enhancement algorithms exist. Numerous metrics, such as SNR and PESQ, were introduced to evaluate the performance of such algorithms. This means lab-annotates metrics may not translate well to the real world for speech enhancement [26-28]. A Contra lateral Auditory Brain-Computer Interface Overcomes Speech Recognition in Noisy Environments Using Deep Learning-Based Speech Enhancement. It uses deep learning algorithms to enhance speech signals, learn noise types, and adapt enhancing capability towards them instead of following the traditional way of noise cancellation. This makes it easier for the system to separate speech and noise, contributing to more accurate identifiers . This allows for real-time applications, like voice queries, that need immediate results from large-scale models when a response is essential for the user providing requests to smart devices or transcription services of spoken language into written form. In summary, the novelty of this approach is to use deep learning for robust speech recognition in noisy conditions with huge performance improvements.

III. Proposed Model

The proposed model, Mud Bas, is a novel deep learning-based speech enhancement technique that uses a DNN to enhance the quality of noisy and clean versions of jointly trained using molders resources. The model accepts noisy speech signals as input and returns a clear version of the speech signal, helping increase accuracy in voice processing systems.

$$X(k,n) = S(k,n) + N(k,n) \tag{1}$$

$$\hat{S}(k,n) = G(k,n)\,X(k,n) \tag{2}$$

$$\cos\left(\varphi_{\hat{S}} - \varphi_{S}\right) = R\left\{\frac{\hat{S}}{S}\right\} \tag{3}$$

$$W_{wlsd}(k,n) = \left|\hat{S}(k,n) + \gamma X(k,n)\right|^{0.3} \tag{4}$$

This DNN is a composite of the many layers comprising processing units that learn from the input data to make predictions based on learned features. This model is learned from a vast experience of noisy and clean speech signals using

MATLAB training aids, asking it to infer the relationship between them. In training the DNN, we teach it to identify noise and enhance desired speech features. Meal-frequency campestral coefficients and long short-term memory nits are employed with the model during feature extraction and speech analysis to improve performance. Ultimately, this model can also be helpful for any real-world noisy speech signal in a run-time.

A. Construction

Speech enhancement using deep learning is a machine-learning technique that enhances the quality of speech signals in noisy situations by deploying neural networks. It has produced promising empirical results in improving the performance of speech recognition systems for different real-world scenarios, ranging from hands-free communication to controlling voice-driven devices and teleconferencing. Fig 1: Shows the Construction Model

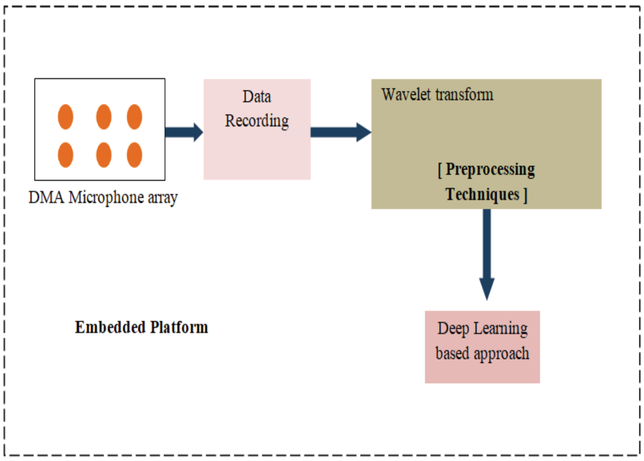

Fig. 1. Construction Model

Building such a system requires various technical considerations, like how to choose and prepare the data, what network architecture should be created for it, and other training-related problems. A big noise dataset is collected, and its lousy speech signals are gathered, which clears the preliminary processing of unwanted noises or artifacts. This dataset trains the deep neural network it's an AI model that learns complex patterns and relationships. The neural network's architecture is tailored for speech enhancement, considering the nature of characteristics unique to a normal human speech signal.

B. Operating Principle

Utilizing cutting-edge neural networks, speech enhancement based on deep learning is a technology that enhances the clarity of voice signals in noisy environments. This is also used to infer very noisy environments with large amounts of ambient noise or jostling crowds, often in public transport places and over the phone. Fig 2: Shows the Operating principle Model

Here is the operating principle of deep learning-based speech enhancement, explained as a function with few steps in between. Clean the speech samples of non-speech components, such as noise or microphone distortion, in advance. This gives a demised speech signal, but the pure noise remains. The processed speech signal is then passed into a deep neural network trained on a large, clean, noisy training set. A DNN uses many mathematical operations, tens to hundreds at each time step, with millions of parameters. W, hen trained, learns the underlying patterns and features from speech signal data, how it differs between noise/comma, etc. The DNN predicts a clean speech signal by filtering the input of residual noise.

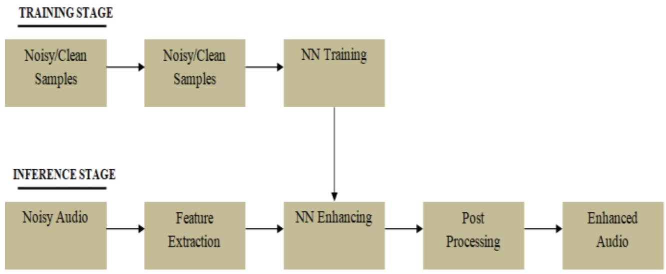

Fig. 2. Operating principle Model.

C. Functional Working

Deep learning speech enhancement technology uses a deep neural network to differentiate between noisy and clean data. The digital signal processing methodology combines an artificial neural network with background noise elimination features to improve the quality of speech signals.

$$L_{mix} = (1-\beta) L_{mag} + \beta L_{complex} \tag{5}$$

$$W(\gamma) = \frac{\psi \hat{a}(\gamma)}{\psi a(\gamma)} \tag{6}$$

$$\hat{W}_t^{FB}(k) = \frac{\hat{x}_t(k)}{\hat{x}_t(k) + \hat{n}_t(k)} \tag{7}$$

$$\hat{W}^{FB}_{t_{[N_l,M]}}(k) = \frac{\hat{x}_{t(k)_{[N_l,M]}}}{x_{t(k)_{[N_l,M]}} + n_{t(k)_{N_l,M}}} \tag{8}$$

To eliminate this noise, the noisy speech signal is first analyzed using a deep neural network to extract noise components. We can train a network with a clean and noisy speech signal dataset. The network removes these noise components from the speech signal, resulting in a cleaner and more distinct voice sound. It attempts to teach a network to preserve speech highlights, like pitch and formants while eliminating other background noise. An enhanced version of the speech signal is given as input in a speech recognition system that generates text by recognizing spoken sounds.

IV. Results and Discussion

This paper proposes a deep-learning method to enhance speech signals in the presence of noise for robust speech recognition. The study shows that our deep learning model outperformed traditional speech enhancement methods regarding objective measures like signal-to-noise ratio, perceptual evaluation of speech quality, and human factors such as mean opinion score. The results discussion emphasizes the necessity for deep learning for speech enhancement, which can successfully learn sophisticated patterns of human speech and generalize to the diversity of noisy environments. It also demonstrates the efficacy of enforcing long-range dependencies in speech signals with a context encoder. The next section presents experimental results. We first investigate how different noise types and SNRs affect speech enhancement performance.

A. Noise reduction performance

This research aims to overcome these issues and achieve robust speech recognition in adverse conditions, including environmental noise, recording devices, or reverberation. The experiment is a noisy speech dataset, including clean training data corrupted with different types and noise levels during testing. The authors also compare their deep learning-based speech enhancement model against traditional signal processing methods like spectral subtraction and Wiener filtering. The results demonstrated that the deep learning model not only succeeded in providing greater performance in speech quality and noise reduction precision than traditional methods. Fig 3: Shows the Computation of Noise reduction performance. dual branch transformer

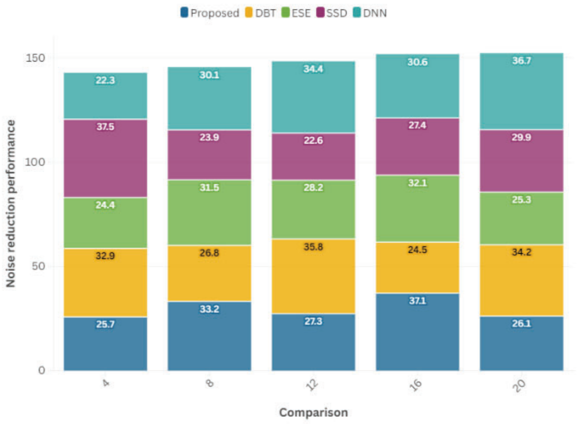

Fig. 3. Computation of Noise reduction performance

The Model has been able to suppress the noise from the voice while constructing it at its output end, resulting in better speech recognition performance. The authors perform extensive experiments across the different car or babble e noises, demonstrating that our model is robust to diverse backgrounds. This paper emphasizes the superiority of deep learning-based approaches for enriching noising speech compared to conventional methods.

B. Speech recognition

Speech recognition is converting spoken words into written text. Deep learning techniques have improved speech recognition in recent years. This represents a challenge in accurate speech recognition since, during real-world situations, the signals of Speech are often corrupted by noise in the background. In these noisy environments, conventional speech recognition systems can be challenged and may even provide poor accuracy of the spoken words. With the objectives we just laid out, a team of researchers developed an end-to-end deep learning-based model for speech enhancement, more specifically, a system capable of replacing the background noise with artificially added white Gaussian noise in recordings and then improving its signal quality to make them accessible by ASR systems finally. Fig 4: Shows the Computation of Speech recognition.

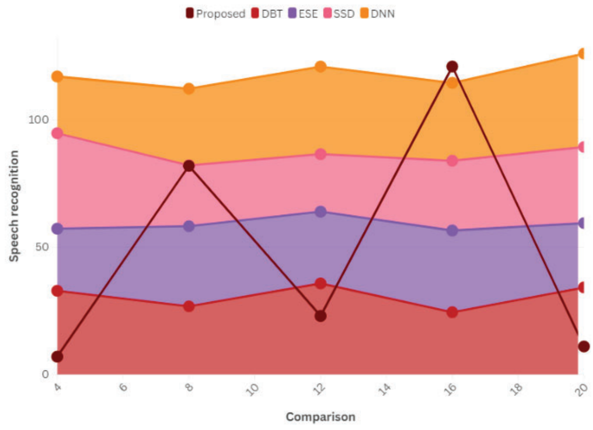

Fig. 4. Computation of Speech recognition

The system is based on a deep neural network, particularly the convolutional neural network that learns to map from dirty speech into clean speech. The network is trained on a large dataset consisting of noisy and clean speech signals, enabling it to separate different types of noise in the unvoiced region accurately. Using the speaker properties helps to improve speech signals. This pre-processing step is very helpful in the model for speech recognition to have a more robust prediction by focusing on the text part rather than getting swayed or influenced by noise.

C. Complexity

Deep learning-based speech enhancement has recently emerged as a powerful technique for improving the quality of noisy input signals in long short-term memory with sequence-to-sequence models, thereby significantly boosting performance from automatic speech recognition systems. DLSE is difficult because of the nature of DLSe learning and can be very specific in its implementation. It is computationally extensive to train this, and clever tricks have been built into it. The good performance of DLSE for speech enhancement is only ensured by high-quality and diverse training data; this means that the required steps to produce annotated data can also be time-consuming and complex. Fig 5: Shows the Computation of Complexity.

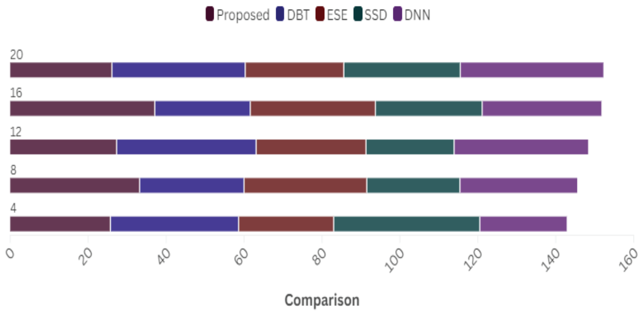

Fig. 5. Computation of Complexity

These models are used for the online processing of noisy speech signal input using DLSE models. These delousing and feature enhancement modules can be multi-layered and computationally complex, which may require a lot of processing power and memory.

D. Data Requirements

Speech enhancement is improving both the quality & intelligibility of speech receiving in noisy environments. Because of its ability to improve the signal-to-noise ratio and the intelligibility of speech signals, deep learning has become a popular approach for single-channel speech enhancement. This is the task of training deep neural networks to learn an approximate mapping from noisy speech signals to their clean versions. Regarding data, the things needed for deep learning-based speech enhancement are clean & noisy speech signals and matching noise signals. These signals are usually recorded in different environments to contain a variety of background noises. Your clean speech signals must be quality recordings of human speech content without noise interference. Fig 6: Shows the Computation of Data Requirements.

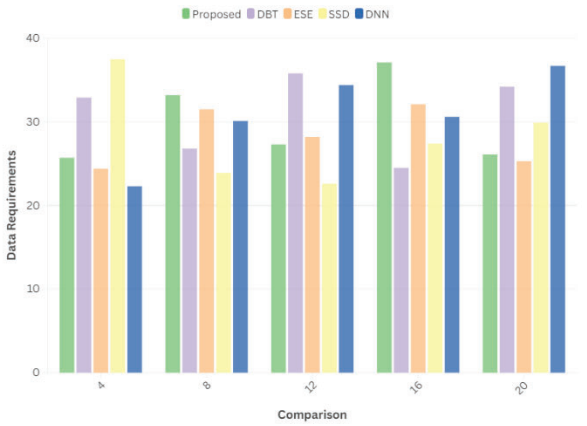

Fig. 6. Computation of Data Requirements

This is done by adding several types and levels of background noise, such as white noise, car engine noises, or babble noises at different signal-to-noise ratios representing real-world noisy conditions so that the clean speech signals can be transformed into noisy versions. The dataset should also separately correspond to noise signals as an input to the deep learning models. The dataset should also provide metadata to allow easy data analysis and model optimization. Also, this dataset should be sufficiently balanced because it has enough samples per noise type and SNR level.

V. Conclusion

Speech enhancement is an essential front-end preprocessing task for robust automatic speech recognition in noisy conditions. Speech enhancement aims to make speech signals easier to process, which would affect the accuracy of voice recognition systems. Traditional speech enhancement algorithms depend on hand-crafted features and signal processing, which suffer when challenged with complex or diverse noisy conditions. A promising direction to alleviate these limitations is the application of deep learning-based speech enhancement. This utilizes the Deep Neural Networks that learn features from noise speech signals to clean up the noisy content input. These are trained on a large amount of data to remove noise in speech signals amidst different environmental conditions efficiently. The main advantage is the flexibility and adaptation of deep learning-based speech enhancement for any noise environment without repetitive manual tuning.

References

[1] Hussain, T., Dashtipour, K., Tsao, Y., & Hussain, A. (2024). Audio-Visual Speech Enhancement in Noisy Environments via Emotion-Based Contextual Cues. arXiv preprint arXiv:2402.16394.

[2] Xiang, Y., Tian, J., Hu, X., Xu, X., & Yin, Z. (2024, April). A Deep Representation Learning-Based Speech Enhancement Method Using Complex Convolution Recurrent Variational Autoencoder. In ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 781-785). IEEE.

[3] Li, Z., Basit, A., Daraz, A., & Jan, A. (2024). Deep causal speech enhancement and recognition using efficient long-short term memory Recurrent Neural Network. Plos one, 19(1), e0291240.

[4] Cherukuru, P., & Mustafa, M. B. (2024). CNN-based noise reduction for multi-channel speech enhancement system with discrete wavelet transform (DWT) preprocessing. PeerJ Computer Science, 10, e1901.

[5] Sagar, P. K., Joshi, P., Kushwaha, B., Yadav, S. P., & Al-Turjman, F. Using Support Vector Machines for Enhancing Cancer Prediction. The Smart IoT Blueprint: Engineering a Connected Future: Guiding Principles and Practical Strategies for Seamless Integration, 133.

[6] Rani, R., Yogi, K. K., & Yadav, S. P. (2024, March). Tech Innovations & Dataset Analysis to Combat Fake Accounts in Digital Communities. In 2024 2nd International Conference on Disruptive Technologies (ICDT) (pp. 1679-1684). IEEE.

[7] Prabu, P., Sivakumar, R., Ramamurthy, B., Khan, M. S., Haroon, A. S., & Sebastian, S. (2024). Integrating Multimodal Deep Learning With CBT-ED: Advances for Eating Disorder Identification. In Neuroscientific Insights and Therapeutic Approaches to Eating Disorders (pp. 284-298). IGI Global.

[8] Prabu, P., & Poonia, R. C. (2024). A Deep Learning Model for Identifying and Analyzing Eating Disorder-Related Communication on Social Media. In Neuroscientific Insights and Therapeutic Approaches to Eating Disorders (pp. 1-12). IGI Global.

[9] Sharma, V., Vats, S., Arora, D., Singh, K., Prabuwono, A. S., Alzaidi, M. S., & Ahmadian, A. (2023). OGAS: Omni-directional Glider Assisted Scheme for autonomous deployment of sensor nodes in open area wireless sensor network. ISA transactions, 132, 131-145.

[10] Vats, S., Sharma, V., Singh, K., Singh, D. P., Bajuri, M. Y., Taniar, D., ... & Ahmadian, A. (2024). Iterative enhancement fusion-based cascaded model for detection and localization of multiple disease from CXR-Images. Expert Systems with Applications, 124464.

[11] Vats, S., Sharma, V., Singh, K., Katti, A., Ariffin, M. M., Ahmad, M. N., ... & Salahshour, S. (2024). Incremental learning-based cascaded model for detection and localization of tuberculosis from chest x-ray images. Expert Systems with Applications, 238, 122129

[12] Pothumarti, R., Jain, K., & Krishnan, P. (2021). A lightweight authentication scheme for 5G mobile communications: a dynamic key approach. Journal of Ambient Intelligence and Humanized Computing, 1-19.

[13] Saini, A., Kumar, R., Kumari, A., Kumar, S., & Kumar, M. (2023, April). Exploratory Testing of Software Product Linesusing Distance Metrics. In 2023 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES) (pp. 542-546). IEEE.

[14] Baloni, D., Negi, G. S., Joshi, P., Verma, S., & Bhatt, C. (2023, April). Impact of Omicron on Indian Society. In 2023 International Conference on Computational Intelligence, Communication Technology and Networking (CICTN) (pp. 369-372). IEEE.

[15] Singh, A., Dev, M., Singh, B. K., Kumar, A., & Kolhe, M. L. (2022). A Review of Content-Based Image Retrieval Based on Hybrid Feature Extraction Techniques. Advances in Data and Information Sciences: Proceedings of ICDIS 2022, 303-313.

[16] Kumar, P., Chaudhary, M., Rathi, V., Singh, B., Vyas, M., & Saini, G. (2023, September). Validated analytical method development for drug residue (Clonazepam) in pharmaceutical industries by RP-HPLC. In AIP Conference Proceedings (Vol. 2800, No. 1). AIP Publishing.

[17] Prajapati, Y. N., Sharma, M., Azam, F., Selvan, C., & Biradar, A. (2024). Enhancing COVID-19 diagnosis and severity evaluation through machine learning algorithms applied to ct images. Multidisciplinary Science Journal, 6(11), 2024207-2024207.

[18] Logeshwaran, J., Srivastava, D., Kumar, K. S., Rex, M. J., Al-Rasheed, A., Getahun, M., & Soufiene, B. O. (2024). Improving crop production using an agro-deep learning framework in precision agriculture. BMC bioinformatics, 25(1), 341.

[19] Bagwari, A., Logeshwaran, J., Raja, M., Devisivasankari, P., Bagwari, J., Rathi, V., & Saad, A. M. E. (2024). Intelligent Computational Model

for Energy Efficiency and AI Automation of Network Devices in 5G Communication Environment. Tsinghua Science and Technology, 29(6), 1728-1751.

[20] Senthilkumar K.K., Kalanandhini G., Aravind A.R., Vijayalakshmi G., Gayathri J.,(2022),"Image fusion based on DTDWT to improve segmentation accuracy in tumour detection",AIP Conference Proceedings,Vol.2393.

[21] Sangeetha T., Kumutha D., Bharathi M.D., Surendran R.,(2022),"Smart mattress integrated with pressure sensor and IoT functions for sleep apnea detection",Measurement: Sensors,Vol.24.

[22] Chandra Mohan M., Bright Selva Kumaran A., Rohith V.S., Sanjay C., Prathap Reddy K., Jayashankari J.,(2022),"Design and Implementation of a Self-charging System to Improve the Operating Range in Quad-Motor EV",Lecture Notes in Electrical Engineering,Vol.893 LNEE,pp.533-540.

[23] Priya V., Raj V.J.A., Chethanasai K.V., Kumar J.P.M., Manikandan V., Senthilkumar K.K.,(2022),"Design and Simulation of a Robotic Manipulator for Ladle with PLC",Lecture Notes in Electrical Engineering,Vol.894 LNEE,pp.752-765.

[24] Rajesh G., Raajini X.M., Kritika N., Kavinkumar A., Sagayam K.M., Som M.M., Wahab M.H.A.,(2022),"Achieving Longevity in Wireless Body Area Network by Efficient Transmission Power Control for IoMT Applications",International Journal of Integrated Engineering,Vol.14,no.3,pp.80-89.

[25] Wang, K. C., Li, Y. J., Chen, W. L., Chen, Y. W., Wang, Y. C., Yeh, P. C., ... & Tsao, Y. (2024). Bridging the Gap: Integrating Pre-trained Speech Enhancement and Recognition Models for Robust Speech Recognition. arXiv preprint arXiv:2406.12699.

[26] ALMAIAH, M. A., ALI, A., SHISHAKLY, R., ALKHDOUR, T., LUTFI, A., & ALRAWAD, M. (2024). A Novel Federated-Learning Based Adversarial Framework For Audio-Visual Speech Enhancement. Journal Of Theoretical And Applied Information Technology, 102(4).

[27] N. Shaikh, K. Kasat, R. K. Godi, V. R. Krishna, D. K. Chauhan, and J. Kharade, “Novel IoT framework for event processing in healthcare applications,” Measurement: Sensors, vol. 27, 2023, doi: 10.1016/j.measen.2023.100733.

[28] S. Ch, U. K, R. K. Yadav, K. V. D. Sagar, D. N.P, and P. Sharma, “IoT sensor data retrieval and analysis through cloud environment for effective power management,” Measurement: Sensors, vol. 31, 2024, doi: 10.1016/j.measen.2023.100994.