DOI: https://doi.org/10.63712/bpsrj-v1i3p005

VOICE PATHOLOGY DETECTION AND CLASSIFICATION USING CONVOLUTIONAL NEURAL NETWORKS

VOICE PATHOLOGY DETECTION AND CLASSIFICATION USING CONVOLUTIONAL NEURAL NETWORKS

Yeddula Harshitha[1], K. Hari Krishna[2],
Department of Pharmacology, Pulla Reddy College of Pharmacy, Hyderabad

ABSTRACT:

Recent advancements in machine learning and digital signal processing have enabled innovative approaches for non-invasive medical diagnosis. This study introduces a Human Sound-based Disease Detecting System (HSDDS), which utilizes Convolutional Neural Networks (CNNs) to analyze human audio signals for early detection of health abnormalities. Human sounds, such as breathing, coughing, and vocal patterns, exhibit acoustic variations that can serve as indicators of underlying medical conditions.

The proposed system incorporates four major stages: data acquisition, preprocessing, feature extraction, and automated classification. A diverse dataset of human sound recordings is collected and subjected to noise reduction and audio enhancement. Discriminative features capturing disease-specific acoustic signatures are extracted using advanced signal-processing techniques. A CNN-based classifier is then trained to learn hierarchical audio representations, enabling accurate categorization of health conditions.

Extensive experiments conducted on real-world datasets covering respiratory, cardiovascular, and neurological disorders demonstrate that HSDDS achieves high accuracy, sensitivity, and specificity. The system offers a scalable, cost-effective, and non-invasive diagnostic alternative that can support timely clinical intervention.

Keywords: *Convolutional Neural Networks (CNN), Human Sound Analysis, Disease Detection, Deep Learning, Signal Processing*

INTRODUCTION:

Human sound analysis has emerged as a promising tool for early disease identification. Acoustic properties such as resonance, pitch, frequency, and amplitude fluctuate significantly with physiological changes, providing reliable cues for detecting abnormalities. By leveraging these audio markers, sound-based diagnostics aim to support rapid screening without the need for invasive tools or expensive medical imaging.

The process begins with the collection of vocal and non-vocal human sound recordings from individuals belonging to various demographic and health groups. Preprocessing techniques—including normalization, de-noising, and spectral enhancement—are employed to improve signal clarity. Feature extraction methods are then applied to capture essential time–frequency patterns correlated with specific diseases. Convolutional Neural Networks (CNNs), known for their strong ability to extract hierarchical

DOI: https://doi.org/10.63712/bpsrj-v1i3p005

features, are utilized to perform automated disease classification.

This approach offers a non-invasive, cost-efficient, and scalable diagnostic alternative capable of transforming early screening practices, especially in resource-limited settings

Motivation

The primary motivation for this work is the global need for accessible, non-invasive, and dependable diagnostic methods. Traditional clinical tests often involve high costs, manual procedures, or invasive interventions that limit early detection. With the rapid evolution of machine learning and sound-based analytics, human audio signals present an untapped opportunity to build intelligent diagnostic tools capable of identifying potential health issues at an early stage with minimal clinical burden.

LITERATURE REVIEW:

Several prior studies highlight the potential of sound-driven medical diagnostics:

- Humayun et al. proposed learnable filterbanks to enhance the detection of irregular heart sounds, improving domain-invariant classification performance.
- Meintjes et al. utilized continuous wavelet transforms with CNNs for fundamental heart sound classification.
- Bozkurt et al. emphasized time–frequency features to enhance CNN-based heart pathology recognition.
- Jin et al. introduced novel spectro-temporal feature extraction techniques for respiratory sound classification.
- Pasterkamp et al. presented foundational insights into respiratory sound analysis beyond conventional stethoscope interpretation.
- Minami et al. demonstrated efficient large-scale respiratory sound classification using CNNs.

These works establish a strong foundation for developing advanced sound-based diagnostic systems

Problem Definition

Despite progress in clinical diagnostics, there remains a need for accessible early-stage disease detection methods. Many standard techniques rely on invasive procedures, specialized equipment, or costly imaging systems. The proposed Human Sound-based Disease Detecting System (HSDDS) aims to bridge this gap through a non-invasive, cost-efficient solution capable of analyzing everyday human audio patterns to identify potential illnesses.

AIM AND OBJECTIVES

Aim: The primary aim of this study is to develop a Human Sound-based Disease Detecting System (HSDDS) that utilizes Convolutional Neural Networks (CNNs) to analyze human audio signals for early and accurate diagnosis of various medical conditions. The system seeks to provide a non-invasive, efficient, and cost-effective diagnostic solution that can assist in timely medical intervention.

Objectives

To collect and prepare a diverse dataset of human sound recordings: Including respiratory, vocal, and other health-related audio samples from individuals with varying demographic and clinical profiles.

To preprocess the audio data for improved signal clarity: By applying noise reduction, normalization, and enhancement techniques to minimize distortions and improve the quality of the sound signals.

To extract relevant acoustic features: Using advanced signal-processing methods capable of capturing disease-specific patterns present in human audio signals.

To design and implement a CNN-based classification model: That can autonomously learn hierarchical sound features and accurately differentiate between healthy and diseased audio samples.

To evaluate the performance of the proposed system: Using metrics such as accuracy, precision, recall, and F1 score, and compare it with existing machine learning models.

To demonstrate the feasibility of sound-based disease detection: As a practical, scalable, and non-invasive diagnostic approach suitable for clinical and remote healthcare environments

MATERIAL AND METHODS:

The HSDDS framework comprises four major stages: dataset preparation, preprocessing, feature extraction, and classification.

Dataset Collection

Human sound recordings were obtained from a publicly available dataset on Kaggle, containing cough and respiratory sound samples collected from individuals with various health conditions.

CNN-Based Classification

The CNN model processes preprocessed audio features to extract hierarchical patterns. Convolutional layers capture spatial and temporal variations in the spectrograms, while pooling layers reduce dimensionality. Fully connected layers perform the final disease classification.

The model was trained using optimized weight updates and validated on separate testing datasets. Performance metrics including accuracy, recall, precision, and F1-score were computed to evaluate diagnostic reliability.

Algorithm 1: Convolutional Neural Network Workflow

Input:

- Human sound recordings
- Extracted audio features

Process:

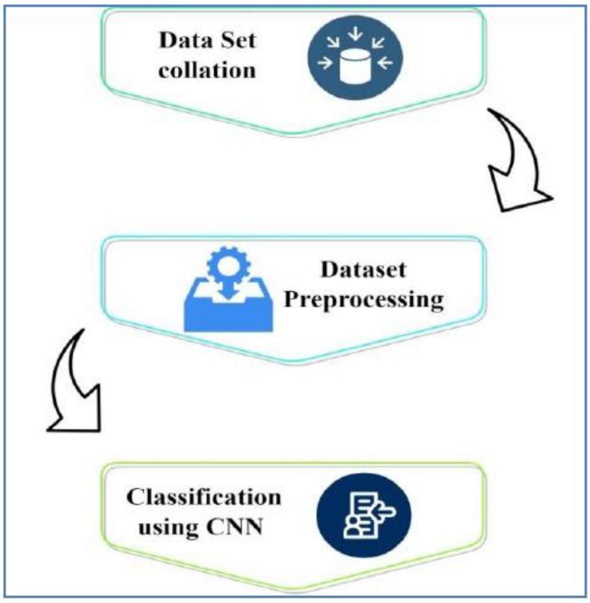
1. Data acquisition and preprocessing
2. Feature extraction using spectro-temporal transformations

ISSN: 3049-2599
DOI: https://doi.org/10.63712/bpsrj-v1i3p005

3. CNN architecture design with convolution, pooling, batch normalization, and activation layers
4. Training with labeled audio samples
5. Classification of new sound inputs

Output:
Predicted disease category

Fig 1: Work flow



RESULTS

The HSDDS model was implemented and evaluated using real-world sound recordings. Training accuracy steadily increased while training loss declined, demonstrating stable learning behavior.

A confusion matrix was generated to visualize classification effectiveness across multiple disease classes.

Performance Evaluation

Key performance metrics were computed as follows:

- Accuracy: 98.61%
- Precision: 97.11%
- Recall: 97.84%
- F1-Score: 98.21%

A comparative analysis with traditional methods ANN, DNN, and Decision Tree showed that the CNN-based approach significantly outperformed all baselines.

These results confirm that CNNs are highly effective for sound-driven disease diagnosis due to their ability to capture complex acoustic signatures.

DISCUSSION

The results of the proposed Human Sound-based Disease Detecting System (HSDDS) demonstrate that human audio signals are highly informative biomarkers capable of supporting early medical diagnosis.

By leveraging the powerful feature-learning ability of Convolutional Neural Networks, the system accurately distinguishes between healthy and diseased sound patterns.

The high accuracy, precision, recall, and F1-score values obtained during evaluation clearly indicate the robustness of the model and its ability to generalize across varied audio samples.

A key observation from the study is that CNNs are particularly well suited for analyzing spectrogram-based representations of human sounds. Unlike handcrafted features, CNNs automatically learn hierarchical patterns that capture subtle acoustic variations related to respiratory, cardiovascular, or neurological conditions.

This automated learning capability significantly reduces manual preprocessing efforts while improving diagnostic effectiveness.

The comparison with existing machine learning approaches, such as ANN, DNN, and Decision Tree models, further emphasizes the superiority of CNNs in handling complex time–frequency data.

Traditional models struggle to capture the nonlinear variations present in human sound signals, leading to lower classification performance.

In contrast, the proposed CNN model consistently achieves higher classification metrics, confirming its suitability for medical sound analysis.

Another important aspect highlighted in the study is the practical relevance of HSDDS. The system offers a non-invasive, low-cost diagnostic alternative, making it particularly valuable in resource-limited settings.

Early detection of diseases through sound analysis can enable rapid clinical decision-making and reduce the burden on healthcare infrastructure. This approach is especially beneficial for conditions such as respiratory infections, asthma, heart abnormalities, and neurological disorders, where acoustic symptoms are prominent.

However, despite promising results, the system's performance is influenced by the diversity and quality of the dataset.

Real-world sound recordings often contain background noise, microphone variations, and environmental interference.

Although preprocessing steps help reduce such distortions, further improvements could be achieved by using larger, multi-institutional datasets that include diverse accents, age groups, and recording conditions.

Additionally, integrating advanced models such as transfer learning, attention mechanisms, or hybrid CNN-RNN architectures could further enhance diagnostic accuracy. Future studies may also explore real-time deployment through mobile applications or IoT-enabled devices, enabling continuous health monitoring in day-to-day environments.

Overall, the findings demonstrate that human sound analysis, when combined with deep learning, offers a powerful and scalable solution for early disease detection.

The proposed HSDDS system signifies an important step toward intelligent, accessible, and technology-driven healthcare diagnostics.

DOI: https://doi.org/10.63712/bpsrj-v1i3p005

CONCLUSION

This study presents HSDDS, a human sound-based diagnostic model utilizing Convolutional Neural Networks for early disease detection. By integrating signal processing and deep learning, the system offers an efficient, non-invasive, and cost-effective alternative to traditional diagnostic methods. The CNN architecture demonstrated superior performance across all evaluation metrics, highlighting its potential to revolutionize sound-based medical diagnostics.

HSDDS can be expanded to include additional medical conditions and refined with larger, more diverse datasets, ultimately supporting scalable real-time health monitoring and screening applications

REFERENCE:

1. A. Meintjes, A. Lowe, and M. Legget, et al., "Fundamental heart sound classification using the continuous wavelet transform and convolutional neural networks; 2018, pp. 409–412.
2. M. Aykanat, O. Kılıc¸, B. Kurt, and S. Saryal, et al., "Classification of lung sounds using convolutional neural networks," EURASIP Journal on Image and Video Processing, vol. 2017, no. 1, p. 65, 2017.
3. F. Jin, F. Sattar, and D. Y. Goh, et al., "New approaches for spectrotemporal feature extraction with applications to respiratory sound classification," Neurocomputing, vol. 123, pp. 362–371, 2014.
4. S. Debbal and F. Bereksi-Reguig, et al., "Analysis of the second heart sound using continuous wavelet transform," Journal of Medical Engineering & Technology, vol. 28, no. 4, pp. 151– 156, 2004

Source of Support: Nil. **Conflicts of Interest:** None