iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** V **Month of publication:** May 2026

DOI: https://doi.org/10.22214/ijraset.2026.82710

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

Smart Diagnostic Tool for Voice Pathologies Using Audio Signal Processing

Riya Singh[1], Dr. Lochan Jolly[2]
Department of Electronics and Telecommunication Engineering, Thakur College of Engineering and Technology, Mumbai, India

Abstract: ***Voice disorders are a lot more common than people think, and catching them early makes a big difference in treatment outcomes. The problem is that most standard diagnostic methods — like laryngoscopy — are invasive, require hospital equipment, and are not practical for regular screening. This paper presents a simple, non-invasive diagnostic tool built on LabVIEW that uses basic audio signal processing to detect voice pathologies in real time. The system records voice samples from both healthy individuals and patients with diagnosed vocal conditions, then analyses them using two straightforward techniques: amplitude-time waveform analysis and power spectral density (PSD). Healthy voices showed smooth, regular waveforms with clear harmonic patterns in the frequency domain, while pathological voices showed irregular waveforms and distorted spectra. Three acoustic markers were also extracted — jitter (0.0237), shimmer (0.1442), and harmonic-to-noise ratio (13.56 dB) — all of which fell outside the normal clinical range for pathological samples. The results confirm that this tool can clearly tell apart healthy and disordered voices without any invasive procedure. It is designed to be affordable, easy to use, and practically deployable in clinics, research labs, or even remote healthcare setups.***

Keywords: Voice Pathology Detection, LabVIEW, Audio Signal Processing, Power Spectral Density, Amplitude-Time Analysis, Dysphonia, Non-Invasive Diagnosis, Vocal Fold Analysis.

I. INTRODUCTION

Voice disorders are more widespread than most people realize. Conditions like dysphonia, vocal nodules, and laryngeal cancer do not just affect how someone sounds — they can seriously impact a person's ability to communicate and, if left undetected, may lead to severe health complications. The usual way to diagnose these conditions is through procedures like laryngoscopy, which is invasive, uncomfortable, and requires specialized hospital equipment. That makes it completely impractical for regular screening or use in resource-limited settings.

Audio signal processing offers a much simpler alternative. Instead of putting a camera down someone's throat, you just record their voice and analyse it. Platforms like LabVIEW are particularly well suited for this because they come with built-in tools for real-time signal acquisition, digital filtering, and spectral analysis — no need to write complex code from scratch [1][2].

This paper describes a LabVIEW-based tool that looks at voice recordings from two angles: the time domain (how the signal behaves over time) and the frequency domain (what frequencies are present). By comparing amplitude-time waveforms and power spectral density plots side by side, the system gives clinicians a clear visual picture of what the vocal folds are doing — without any invasive procedure. The tool is designed for clinical settings, research labs, and remote healthcare environments where expensive diagnostic equipment may not be available.

The main contributions of this work are: (i) a complete, real-time signal processing pipeline built entirely in LabVIEW; (ii) extraction and interpretation of three acoustic markers — jitter, shimmer, and harmonic-to-noise ratio; and (iii) a side-by-side visual interface that lets clinicians compare healthy and pathological voice patterns easily.

II. LITERATURE SURVEY

Research on voice disorder detection has been going on for over two decades, and the field has come a long way. Early methods relied on clinical tools like videolaryngoscopy and the Computerized Speech Lab (CSL). These work well but are expensive, require trained personnel, and are not scalable for widespread use.

A lot of research has focused on acoustic features like jitter, shimmer, harmonic-to-noise ratio (HNR), and formant frequencies as indicators of vocal fold health. These features show measurable differences between healthy and disordered voices, which makes them useful for non-invasive diagnosis. The challenge is that extracting these features accurately often requires specialized software and expertise, which limits their practical use in everyday clinical settings.

©IJRASET: All Rights are Reserved

Gurdak et al. [1] showed that LabVIEW could be used for computer-aided diagnosis of laryngeal disorders, combining signal acquisition and classification in a single environment. Amrutha et al. [2] confirmed that LabVIEW is capable of real-time acoustic characterization of speech. Martinek [3] took this further by integrating National Instruments microphones, demonstrating that clean signal acquisition is possible even in noisy environments.

More recently, machine learning approaches using support vector machines, random forests, and neural networks have been applied to voice classification. These can achieve high accuracy but require large labelled datasets that are difficult to collect and annotate, and they are computationally heavy to run in real time on standard hardware. Banacia [6] and Todorov [7] pointed out that LabVIEW's built-in signal analysis tools offer a practical, interpretable alternative — especially when the goal is real-time diagnostics without a complex classification pipeline. This work builds directly on those ideas.

III. DATASET DESCRIPTION

Voice samples were collected from two groups: healthy volunteers with no history of voice problems, and patients clinically diagnosed with vocal disorders including in-situ carcinoma and dysphonia. All recordings were done in a quiet, controlled environment using a high-quality condenser microphone to keep background noise to a minimum.

Both sustained vowel sounds (specifically the phoneme /a/) and short segments of continuous speech were recorded to cover different phonation modes. When live recording was not possible, pre-recorded WAV files sampled at 16,000 Hz with 16-bit resolution were used instead. Table I gives a summary of the dataset.

Parameter	Healthy Group	Pathological Group
Recording Type	Live + Pre-recorded WAV	Live + Pre-recorded WAV
Phonation Mode	Sustained /a/, continuous speech	Sustained /a/, continuous speech
Sampling Rate	16,000 Hz	16,000 Hz
Bit Depth	16-bit	16-bit
Diagnosis	None (healthy)	In-situ carcinoma, dysphonia
Environment	Controlled, low-noise	Controlled, low-noise

Table I: Dataset Composition Summary

IV. PROPOSED METHODOLOGY

The system works through a four-stage pipeline: voice acquisition, preprocessing, feature extraction, and output visualization. Every stage runs inside LabVIEW, which acts as the single platform for the entire workflow — from recording the voice to displaying the diagnostic plots.

A. System Architecture

The pipeline has four modules, each built independently so they can be tested and updated separately. The acquisition module records voice through a standard microphone or loads a saved WAV file. The preprocessing module cleans the signal by removing noise and normalizing the amplitude. The feature extraction module then computes the amplitude-time waveform and PSD plot. Finally, the visualization module displays both plots side by side on the LabVIEW front panel so clinicians can compare them directly. Table II lists each module along with the LabVIEW tools used.

Module	Function	LabVIEW Tools
Voice Acquisition	Capture live input or load WAV files	DAQ Assistant, Sound File Read VI
Preprocessing	Noise filtering, amplitude normalization	Digital Filter Design, Normalize VI
Feature Extraction	Amplitude-time and PSD computation	Waveform Graph, FFT Power Spectrum VI
Visualization	Side-by-side diagnostic plot display	Front Panel Graphs, Waveform Chart

Table II: System Architecture Modules

B. Functional Requirements

The system handles both sustained vowel sounds and continuous speech from a live microphone or a pre-recorded file. Preprocessing runs in real time, and the amplitude-time and PSD outputs appear within seconds of receiving the voice input. The front panel layout makes it easy to compare healthy and pathological patterns side by side, with clearly labelled axes and properly scaled plots.

C. Non-Functional Requirements

Results come out within a few seconds of input, which is fast enough for clinical use. The system works reasonably well across different recording environments and microphone types. The interface is kept simple enough that a speech therapist or clinic staff member can use it without knowing anything about signal processing. Future versions of the tool are planned for integration with portable hardware and telemedicine platforms. All voice data is processed locally on the machine — nothing is sent externally without the user's consent.

V. SIGNAL PREPROCESSING

Before any analysis happens, every voice recording goes through a standard preprocessing sequence. This step is important because it ensures the feature extraction stage always works on clean, consistent data — regardless of how or where the recording was made.

A. Noise Reduction

A digital band-pass filter (FIR type) is applied to cut out low-frequency hum from the environment and high-frequency electrical noise from the microphone or equipment. The passband is set from 80 Hz to 8,000 Hz, which covers the full range of clinically useful vocal frequencies while removing everything outside it.

B. Amplitude Normalization

All recordings are scaled to the same peak amplitude before comparison. This removes the effect of different microphone sensitivities or recording volumes, so when two signals are compared, any differences in waveform shape or spectral energy actually reflect differences in the voice itself — not the recording setup.

C. Segmentation

For sustained vowel recordings, a fixed-length segment from the middle of the recording is extracted to avoid the transients at the start and end of phonation. For continuous speech, voiced segments are identified automatically using energy-based endpoint detection, and only those segments are passed to the feature extraction stage.

VI. FEATURE EXTRACTION

Once preprocessing is done, the cleaned signals are analysed in two ways — in the time domain and in the frequency domain — to capture different aspects of how the vocal folds are behaving.

A. Time-Domain Analysis

The amplitude-time waveform shows how the signal's amplitude changes over the duration of the recording. In healthy voices, this waveform looks regular and periodic — the vocal folds open and close in a steady rhythm. In pathological voices, the waveform becomes irregular: you can see sudden drops in amplitude, random peaks, and an overall lack of consistency, which reflects the physical disruption caused by nodules, tumours, or other tissue abnormalities.

B. Frequency-Domain Analysis (PSD)

Power spectral density tells you how the signal's energy is spread across different frequencies. It is computed using the Fast Fourier Transform (FFT). For a discrete voice signal x[n] of length N, the PSD is estimated as:

$$P(fk) = (1/N)\,|X(fk)|^2, \quad k = 0, 1, \ldots, N/2$$

Here, X(fk) is the FFT of the signal at frequency bin fk. A healthy voice shows clear harmonic peaks spaced evenly across the frequency axis, which is exactly what you expect from a regular, periodic vibration source. A pathological voice, on the other hand, shows a messy spectrum — missing harmonics, elevated noise across a wide frequency range, and sometimes unexpected peaks at unusual frequencies, which suggests the vocal fold tension or mass has changed.

C. Acoustic Marker Extraction

Beyond the waveform and PSD plots, three specific acoustic markers are computed to give a numerical measure of vocal fold irregularity. Jitter measures how much the pitch varies from one cycle to the next. Shimmer measures how much the amplitude changes between cycles. HNR (harmonic-to-noise ratio) measures how much of the signal is clean harmonic content versus noise. Table III shows the values extracted from the test samples, the normal clinical ranges, and what the values suggest.

Feature	Extracted Value	Normal Range	What It Indicates
Jitter	0.0237	0.002 – 0.010	Above normal — vocal fold pitch perturbation
Shimmer	0.1442	0.020 – 0.100	Above normal — breathy or weak voice quality
HNR (dB)	13.56 dB	20 – 40 dB	Below normal — noisy or breathy phonation

Table III: Extracted Acoustic Features and Clinical Interpretation

VII. CLASSIFICATION APPROACH

This system does not use a machine learning classifier. The classification is done in two straightforward ways: by checking acoustic marker values against known normal ranges, and by visual inspection of the waveform and PSD plots. This keeps the system fast, transparent, and easy to understand — which matters a lot in a clinical setting.

A. Threshold-Based Assessment

The three extracted markers — jitter, shimmer, and HNR — are each compared against their clinical reference range (shown in Table III). If two or more markers fall outside the normal range, the voice sample is flagged as potentially pathological. Requiring at least two markers to be abnormal helps avoid false alarms from a single noisy reading.

B. Visual Pattern-Based Assessment

The amplitude-time waveform and PSD plot are displayed side by side so the clinician can visually confirm what the numbers are saying. Irregular waveforms and distorted harmonic structures are easy to spot even without signal processing expertise. This combination of numerical flags and visual output gives the clinician two independent lines of evidence before making any clinical judgment.

VIII. EXPERIMENTAL RESULTS

The system was tested on voice recordings from healthy volunteers and patients diagnosed with in-situ carcinoma. The results were consistent across all samples and clearly showed the system can tell the two groups apart.

A. Time-Domain Results

For healthy voices, the amplitude-time waveform was regular and periodic — exactly what you expect when the vocal folds are opening and closing normally. For pathological voices, the waveforms were visibly different: irregular amplitude, sudden dips, and erratic peaks. This kind of pattern makes sense physically — when pathological tissue like a tumour or nodule disrupts the vocal folds, the vibration pattern breaks down.

B. Frequency-Domain Results

Healthy voice PSD plots showed a clear set of harmonic peaks spaced evenly across the frequency axis — a well-organized harmonic structure that is consistent with steady, periodic vocal fold vibration. Pathological voice PSD plots looked very different: harmonics were missing or heavily distorted, the noise floor was elevated across a wide frequency range, and some samples showed unexpected dominant peaks at unusual frequencies — suggesting a shift in fundamental frequency caused by changes in vocal fold mass or stiffness.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

C. Acoustic Marker Results

All three markers extracted from pathological samples were outside normal ranges: jitter at 0.0237 (normal: 0.002–0.010), shimmer at 0.1442 (normal: 0.020–0.100), and HNR at 13.56 dB (normal: 20–40 dB). These numbers line up well with what was observed in the waveforms and PSD plots — all three analytical methods are pointing to the same conclusion.

D. System Performance

Both the amplitude-time plot and the PSD graph were generated within a few seconds of providing the voice input, which confirms the system is fast enough for practical clinical use. Plot quality was consistently clear and well-labelled. The system worked equally well with live microphone input and pre-recorded WAV files, which shows it is robust across different input types.

IX. COMPARATIVE ANALYSIS

Table IV compares the proposed system with conventional clinical voice assessment and machine learning-based approaches across the parameters that matter most for practical deployment.

Parameter	Conventional (Laryngoscopy)	ML-Based Systems	Proposed System
Invasiveness	Invasive	Non-invasive	Non-invasive
Real-Time Output	No	Limited	Yes (few seconds)
Hardware Required	Specialized clinical setup	GPU / server	Standard mic + PC
Dataset Dependency	None	Large labelled dataset	Small sample set
Interpretability	Expert-only	Low (black-box)	High (visual plots)
Cost	High	Moderate to High	Low
Where It Works	Hospital only	Lab / research	Clinic / remote

Table IV: Comparison of Voice Diagnostic Approaches

Table V gives a direct comparison of what was observed in the healthy versus pathological voice recordings across all the features analysed.

Acoustic Feature	Healthy Voice	Pathological Voice
Amplitude Waveform	Regular, periodic oscillations	Irregular, erratic fluctuations
Harmonic Structure	Clear peaks at even intervals	Missing or distorted harmonics
Spectral Noise	Low background noise	Elevated noise across spectrum
Dominant Frequency	Steady fundamental frequency	Shifted or unstable peaks
PSD Pattern	Uniform energy distribution	Abnormal spectral spikes
Jitter	0.002 – 0.010 (normal)	0.0237 (elevated)
Shimmer	0.020 – 0.100 (normal)	0.1442 (elevated)
HNR (dB)	20 – 40 dB (normal)	13.56 dB (low)

Table V: Feature Comparison — Healthy vs. Pathological Voices

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

X. LIMITATIONS

This system works well for what it is designed to do, but there are some honest limitations worth acknowledging. The testing was done on a relatively small set of samples. Before this tool is used more widely, it needs to be validated on a larger and more diverse group of patients with different types of voice disorders and across different age groups and demographics.

The threshold-based classification may struggle with borderline cases where the acoustic markers are only slightly outside the normal range. The system has also been tested mainly on sustained vowel sounds, not on natural conversational speech. Background noise in uncontrolled environments can still affect result quality even after filtering. Most importantly, this tool is meant to support a clinician's judgment — not replace it. Any result flagged by the system must be reviewed by a qualified medical professional before any clinical decision is made.

XI. FUTURE SCOPE

There are several directions where this system can be taken forward. The most practical near-term improvement would be porting the tool to portable hardware — a small embedded system with wireless connectivity — so it can be used for community screening and telemedicine without needing a full PC setup. Adding longitudinal monitoring would allow clinicians to track a patient's vocal health over time and catch slow-progressing disorders earlier. A lightweight rule-based or statistical classifier could be added on top of the existing visual outputs to give an automatic preliminary diagnosis without the overhead of deep learning. Expanding the feature set to include formant frequencies and MFCCs would give a richer picture of the voice. A mobile app interface would make the tool much more accessible for patients and healthcare workers in remote areas.

XII. CONCLUSION

This paper presented a LabVIEW-based tool for detecting voice pathologies using straightforward audio signal processing — no invasive procedures, no expensive equipment, no complex machine learning. The system analyses voice recordings through amplitude-time waveforms and PSD plots, and backs those up with three acoustic markers: jitter, shimmer, and HNR.

Testing on samples from healthy individuals and patients with in-situ carcinoma showed clear, consistent differences between the two groups. All three acoustic markers were outside normal ranges for pathological samples, and this matched what the waveforms and spectral plots were showing visually. The system processes each recording in a matter of seconds, works with both live microphone input and pre-recorded files, and produces output that is easy to read without any technical background in signal processing. The goal of this tool is to give ENT specialists, speech therapists, and general clinicians a practical, affordable first line of assessment — especially in settings where standard diagnostic equipment is not available. With further development, including validation on larger datasets and integration with portable hardware, this system has the potential to become a genuinely useful tool in everyday voice healthcare.

XIII. ACKNOWLEDGEMENT

The authors would like to thank the Department of Electronics and Telecommunication Engineering, Thakur College of Engineering and Technology, Mumbai, for providing the necessary support and resources to carry out this project. The authors also express their sincere gratitude to their guide, Dr. Lochan Jolly, for her continued support, encouragement, and valuable feedback throughout the development of this work.

REFERENCES

[1] D. Gurdak, K. Pancerz, J. Szkola, and J. Warchol, "Computer-Aided Diagnosis of Laryngopathies in the LabVIEW Environment: Exemplary Implementation," in Advances in Intelligent Analysis of Medical Data and Decision Support Systems, R. Kountchev and B. Iantovics, Eds., Studies in Computational Intelligence, vol. 473. Cham, Switzerland: Springer, 2013, pp. 157-167.

[2] A. S. Amrutha, A. S. Nair, C. J. Mathew, J. E. George, and M. George, "Characterization of Speech Signal in LabVIEW Platform," Int. J. Eng. Res. Technol., vol. 3, no. 5, pp. 1-4, 2015.

[3] R. Martinek, "Speech Signal Processing Using Microphones NI-9234 and LabVIEW," in Proc. Int. Conf. Computer Modeling and Simulation, 2018, pp. 112-117.

[4] National Instruments, Signal Analysis -- LabVIEW API Reference, Austin, TX, USA: National Instruments, 2025.

[5] National Instruments, LabVIEW Jitter Analysis Toolkit -- Overview, Austin, TX, USA: National Instruments, 2014.

[6] A. Banacia, "A Learning Module on Sound Analysis Using LabVIEW," in Proc. IEEE Int. Conf. Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM), 2015, pp. 1-5.

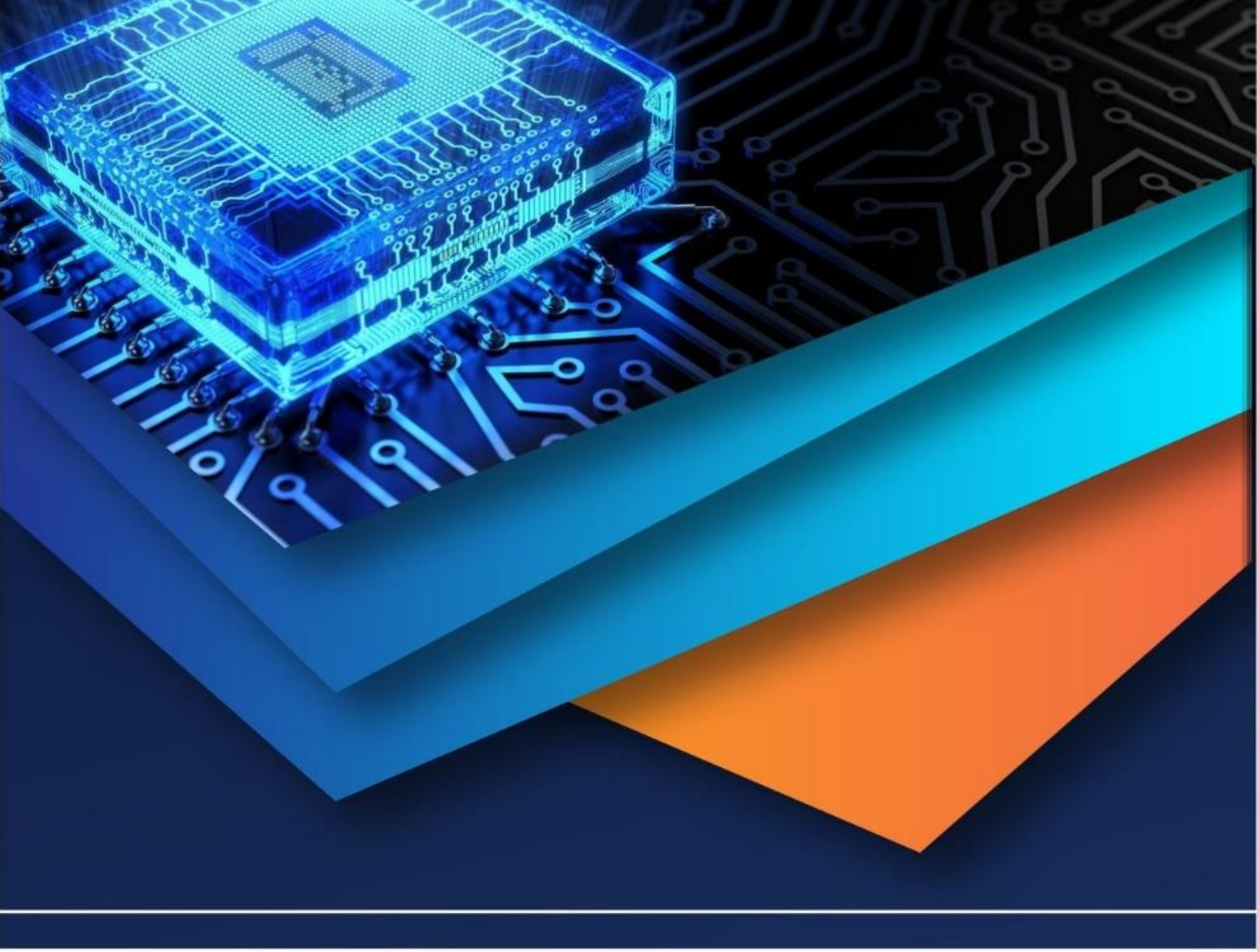
[7] R.-C. V. Todorov, "Audio Signal Processing System Design Using LabVIEW," in Proc. Int. Conf. Electronics, Signal Processing and Communication Systems (ICEST), 2015, pp. 225-229.

[8] S. K. Singla and R. K. Yadav, "Optical Character Recognition Based Speech Synthesis System Using LabVIEW," J. Appl. Res. Technol., vol. 12, no. 5, pp. 843-850, 2014.

[9] National Instruments, LabVIEW Advanced Signal Processing Toolkit User Manual, Austin, TX, USA: National Instruments, 2025.

[10] National Instruments, "Using LabVIEW for Voice Signal Analysis," NI Community Technical Report, Austin, TX, USA: National Instruments, 2008.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)