IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Neuro-Driven Speech Synthesis

[1]Anagha Prakash,[2]Anirudha R Bhat,[3]Gurushankara M,[4]Bharathi Gururaj

Student, Dept of ECE, K S Institute of Technology, Affiliated to VTU, Bangalore, India[1-3]

Professor, Dept of ECE, K S Institute of Technology, Affiliated to VTU, Bangalore, India[4]

Abstract: Speech is a fundamental means of communication that allows individuals to express thoughts, emotions, and ideas. However, millions of people worldwide are unable to communicate verbally due to conditions such as amyotrophic lateral sclerosis (ALS), brainstem stroke, locked-in syndrome, or severe paralysis. Traditional augmentative and alternative communication (AAC) devices, such as eye-tracking systems or text-based interfaces, are often slow, labour intensive, and less expressive. In recent years, advancements in neuroscience, machine learning, and brain-computer interface (BCI) technologies have paved the way for neuro-driven speech synthesis, which holds the promise of restoring communication for individuals with severe speech impairments.

Index Terms - Brain-Computer Interface, Speech Impairment, Human-Computer Interaction, EEG Signal Processing, Deep Learning, Assistive Communication.

I. Introduction

Understanding Neuro-Driven Speech Synthesis

Neuro-driven speech synthesis involves decoding neural signals directly from the brain to generate speech or text. This approach bypasses the need for muscle movements, making it ideal for individuals with complete paralysis. The process typically involves recording brain activity using non-invasive methods such as electroencephalography (EEG) or invasive methods like electrocorticography (ECoG). Neural signals are processed to extract patterns corresponding to speech related activity, which are then decoded into linguistic units or synthesized as spoken words using speech synthesis models.

The field has gained momentum due to rapid advancements in BCI technologies. BCIs serve as a bridge between the brain and external devices, enabling direct communication by interpreting neural activity. For neuro-driven speech synthesis, BCIs leverage advanced algorithms to decode complex neural signals associated with speech perception, production, or imagination. This emerging technology has the potential to revolutionize communication for individuals with disabilities, enhancing their quality of life and autonomy.

Recent Advancements and Challenges: The field of neuro-driven speech synthesis has seen significant progress in recent years. Notable studies have introduced innovative datasets like 'Chisco' for imagined speech decoding and proposed end-to-end frameworks for decoding listened speech. Hybrid deep learning models combining spatial and temporal analysis have improved decoding accuracy. Furthermore, real-time synthesis systems using chronically implanted BCIs have demonstrated the feasibility of online speech generation for patients with ALS.

Despite these advancements, several challenges remain. EEG-based systems face limitations due to low spatial resolution and susceptibility to noise. The variability in neural activity across individuals makes it difficult to develop generalized models. Moreover, achieving real-time decoding and synthesis with high accuracy and low latency is a technical hurdle. Ethical considerations, such as user privacy and the invasiveness of certain techniques, also require careful attention.

Potential Applications: Neuro-driven speech synthesis can have transformative applications beyond restoring communication for individuals with paralysis. It can enhance silent communication technologies, assist in language learning, and provide insights into the neural basis of speech and language. Additionally, the integration of such systems with virtual reality (VR) and augmented reality (AR) could open new possibilities for immersive communication experiences.

II. Literature Survey

Toward Fully-End-to-End Listened Speech Decoding from EEG This research proposes FESDE, a novel end-to-end framework for decoding listened speech directly from EEG signals. Unlike traditional multi-step approaches, FESDE reconstructs speech waveforms without intermediate acoustic feature extraction. The model consists of an EEG module, a speech module, and a connector that maps EEG embeddings to speech representations. The study demonstrates improved efficiency and decoding accuracy, making it a promising development for EEG-based speech synthesis. [1]

Chisco: An EEG-based BCI Dataset for Decoding of Imagined Speech This paper introduces the Chinese Imagined Speech Corpus (Chisco, a large EEG dataset focused on imagined speech decoding. It consists of over 20,000 sentences recorded from healthy adults, with each participant contributing more than 900 minutes of EEG data. The study aims to advance non-invasive Brain-Computer Interfaces (BCIs) by enabling sentence-level semantic reconstruction of imagined speech. It discusses experimental paradigms, data collection, and preprocessing methods, making Chisco a valuable resource for neural language decoding. [2]

EEG Signal Processing for Medical Diagnosis, Healthcare, and Monitoring: A Comprehensive Review This review paper explores EEG signal processing techniques for diagnosing neurological disorders such as epilepsy, Alzheimer's, and autism. It covers preprocessing methods, feature extraction, and classification algorithms, including machine learning and deep learning approaches. The study also highlights challenges like noise removal, data variability, and the need for better feature engineering. It serves as a reference for researchers aiming to enhance EEG-based medical diagnostics.[3]

Exploring the Frontier: Transformer-based Models in EEG Signal Analysis for Brain-Computer Interfaces This review examines the use of Transformer-based deep learning models in EEG signal processing for BCIs. It discusses architectures such as the Temporal Spatial Transformer Network (TSTN) and EEG Conformer, which improve EEG data classification and noise reduction. The paper highlights how attention mechanisms enhance interpretability and real-time processing. It also suggests future research directions for integrating Transformers with BCIs beyond traditional motor imagery applications.[4]

Thought-Controlled Wheelchair Using EEG Acquisition Device This paper presents a non-invasive EEG-based wheelchair control system that enables mobility for individuals with physical disabilities. It describes a system that captures EEG signals using an electrode cap and processes them to generate wheelchair movement commands. The study details signal acquisition, preprocessing, transmission, and motor control mechanisms. It highlights the advantages of non-invasive BCIs and discusses challenges like signal noise and response time.[5]

Continuous and Discrete Decoding of Overt Speech with Non-Invasive EEG This paper examines methods for decoding overt speech using EEG signals, focusing on both continuous and discrete decoding approaches for real-world applications. The study evaluates techniques to extract speech features from EEG in real time, enabling continuous synthesis of speech or discrete classification of words. It highlights the potential of EEG as a reliable modality for speech decoding. It includes flexibility in applications (e.g., continuous speech restoration or command-based systems), real-time performance, and adaptability to noisy environments. The findings contribute to the development of assistive technologies for individuals with speech impairments.[6]

Diff-E: Diffusion-Based Learning for Decoding Imagined Speech EEG This paper introduces 'Diff-E,' a diffusion-based learning model for decoding imagined speech from EEG signals. It utilizes denoising diffusion probabilistic models to enhance signal clarity and decoding performance. 'Diff-E' addresses the stochastic nature of EEG signals by applying advanced probabilistic learning techniques. This model demonstrates superior decoding accuracy compared to traditional methods. It includes enhanced robustness to noise, adaptability to diverse signal types, and potential for real-time applications in BCIs. The study opens new avenues for handling complex EEG data in silent speech communication systems.[7]

Neuro Talk: Towards Voice Reconstruction from EEG during Imagined Speech 'Neuro Talk' is a model designed to reconstruct the user's voice from imagined speech EEG signals. It integrates personalized voice synthesis with neural decoding for silent communication. This approach converts non-invasive EEG signals of imagined speech into synthesized speech resembling the user's voice. The model adapts to unseen words, showcasing its versatility. As an advantage this includes personalized voice reconstruction, applicability to real-world BCIs, and potential for silent communication technologies. The study highlights the promise of combining neural decoding with advanced synthesis techniques for assisting speech-disabled individuals.[8]

III. OBJECTIVES

- To improve communication and standard of living for individuals with speech impairment or vocal issues.
- Creation of affordable and accessible communication devices to help those in need.
- Automation and innovation in the field of healthcare and hospitality.
- Protect privacy and authenticity of individuals using the device.
- Optimization and real time result driven performance for smooth communication.

IV. METHODOLOGY

4.1 Working

The process starts by putting gel-based electrodes on the scalp according to the International 10-20 Electrode Placement System to get the best EEG signal recording. These electrodes record neural activity in the form of electrical signals, which are very weak and prone to noise interference. To improve signal purity, the Bio Amp EXG Pill is applied as a bio signal amplifier prior to transmission of the signals for processing. The EEG signals are thereafter captured for 20 minutes across numerous sessions in an enclosed setting in order to limit artifacts from activity of muscles or environmental noise. The raw data is processed and formatted using Visual Studio Code prior to being forwarded for sophisticated analysis. The EEG data is uploaded onto Google Colab, where preprocessing methods like noise removal, baseline removal, and separation of frequency bands clean the signals.

Buthworth filters are used to remove noise due to blinking and head movement, allowing only significant features to be used for training. The pre-processed data is then converted into structured CSV files, ready for use with machine learning frameworks. The data is divided into training, validation, and test subsets so that a deep learning model—a Recurrent Neural Network (RNN) or a Transformer-based model—can learn to identify EEG signal patterns for certain thoughts. After training, the model predicts target words from brain signals and is loaded onto an Arduino Uno microcontroller, which accepts real-time EEG inputs and produces corresponding speech through a speaker. Feedback mechanisms allow users to fine-tune the system for better accuracy.

The model is tested in real-time in various environments to evaluate its resilience against noise interference, and adaptive learning methods improve its performance. Optimization techniques like hyperparameter tuning and dropout layers avoid overfitting. After successful validation, the system is deployed for practical applications, helping people with severe communication disorders like ALS and Locked-In Syndrome. The next steps will include the incorporation of the system into a wearable EEG headset for seamless use, implementing multilingual features, and creating compatibility with smart home appliances. The final vision is to provide a direct correlation between brain signals and speech to revolutionize assistive communication in individuals with severe speech disabilities.

4.2 Block Diagram

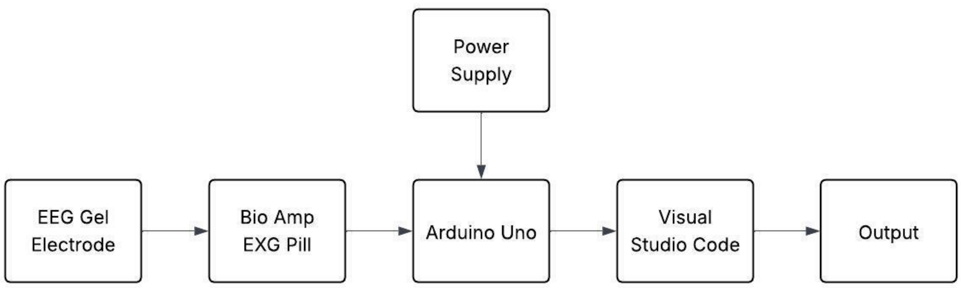

Fig.1 block diagram for Neuro-driven speech synthesis

The above figure shows the working of the project neuro-driven speech synthesis. Here the EEG Gel Electrode is placed on the scalp of the patient. This will take the data from the brain which will be stored in the csv format. Data which we get will be in the integer ranges from 0 to 500 will be mapped as food and integer from 500 till 1000 will be mapped as water. This will take 20 minutes for each case. This collected data is used for predicting either food or water. This prediction is done in Google colab and the end result is sent to the Visual Studio Code, where a prefeed audio for food and water is present which will be heard via the speakers as an output of this project.

V. COMPONENTS

- EEG Gel Electrode

Fig.2 EEG Gel Electrode

The EEG Gel Electrode is a vital part of the brain-computer interface (BCI) system, used to capture the electrical activity of the brain and send it for processing. These electrodes are applied to the scalp to record electroencephalogram (EEG) signals, which are produced by neurons activating in the brain. The usage of gel-based electrodes increases conductivity by minimizing impedance between the skin and the electrode, providing a cleaner and more precise signal. These electrodes are generally constituted of silver-silver chloride (Ag/AgCl) material, renowned for its superior signal transmitting abilities.

- Arduino Uno

Fig.3 Arduino Uno

The Arduino Uno stands as a cornerstone in the realm of microcontroller boards, renowned for its reliability and versatility in electronics and do-it-yourself (DIY) projects alike. Central to its operation is the ATmega328P microcontroller, clocked at a brisk 16 MHz, providing a robust foundation for a myriad of applications. With 14 digital input/output pins, including 6 PWM-capable ones, alongside 6 analog input pins, the Uno boasts adaptability crucial for interfacing with various sensors and devices. Operating at a voltage of 5V, it accommodates a broad spectrum of components. Its onboard memory resources further enhance its capabilities, offering 32KB of flash memory for program storage, 2KB of SRAM for variable storage, and1KB of EEPROM for data retention. The inclusion of a USB interface streamlines both programming and power supply, facilitating seamless connectivity with computers.

- BioAmp EXG Pill

Fig.4 Bio Amp EXG Pill

Bio Amp EXG Pill is an integral component of this project, a miniature high-speed biopotential amplifier for taking electrical signals off of the body of a human. In our system, it has been implemented for filtering and amplifying raw EEG signals gained by the EEG Gel Electrodes located on the head of a user. Because brain signals are weak and very vulnerable to electrical interference from muscle movements, electrical noise, and other external environmental interference, the Bio Amp EXG Pill amplifies the signal and suppresses unwanted interference. This ensures clean, high-quality EEG data is provided to the Arduino Uno to be processed. The pill is small, light, and made for easy use with microcontrollers such as Arduino, making it perfect for real-time collection of brainwave signals.

- VS Code

Fig.5 VS Code

Visual Studio Code (VS Code) is a lightweight yet powerful code editor that is the main software environment for processing EEG signals in this project. It is employed to execute two major Python scripts: collect.py and predict.py, which process the signal acquisition and prediction operations. The collect.py script stores the live EEG signals from Arduino for subsequent processing, whereas predict.py retrieves classified signals from a trained Google Colab hosted machine learning model. VS Code provides instant integration with Python libraries supporting data handling, signal processing, and cloud communications, making for smooth execution of the brain-computer interface (BCI) system. Also, the debugging features, extensions, and terminal integration of the platform make development and debugging easy, a feature that is suited for applying real-time EEG signal interpretation.

- Google Colab

Fig.6 Google Colab

Google Colab is pivotal to the project because it acts as the cloud-based processor that performs training and run of the machine learning model for the classification of EEG signals. Brainwave signals are intricate in nature and are subject to complex interpretation using superior neural networks; thus, Google Colab offers a great platform for easy and powerful training and prediction using deep learning models. The EEG signals picked up through collect.py in VS Code are forwarded to Colab, which runs a pre-trained machine learning model to process them to determine the patterns related to particular thoughts. The predicted results are then transmitted back to VS Code via predict.py, which then transforms them into voice output through the speaker of the laptop. Google Colab's Python library integration with TensorFlow, NumPy, and Pandas provides effective signal analysis, and its GPU acceleration over the cloud acts substantially to enhance the speed of model training and deployment. This erases hardware constraints, and computationally demanding EEG signal processing can be performed in real-time.

VI. RESULTS

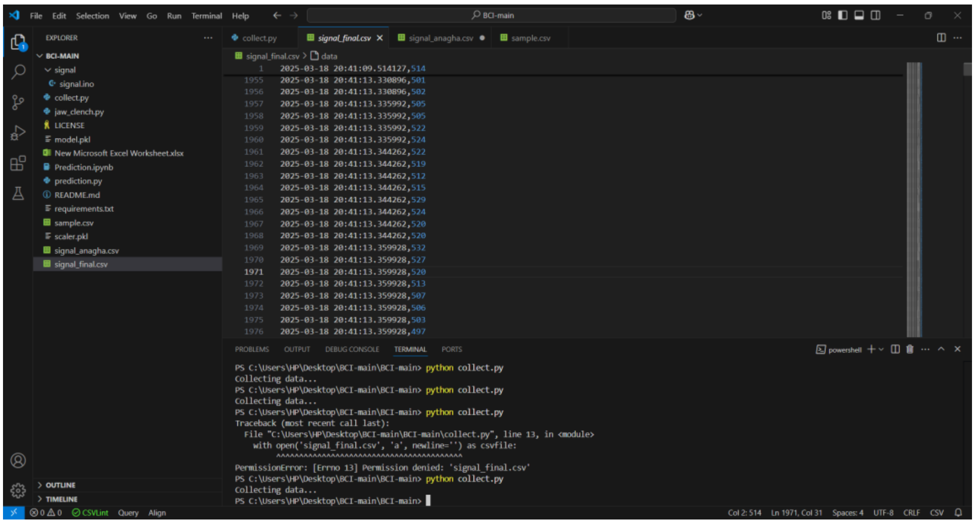

Fig.7 As it can be seen in the above picture, it takes about 20 minutes to collect 4 sets of data. After setting the command prediction.py the data collection occurs.

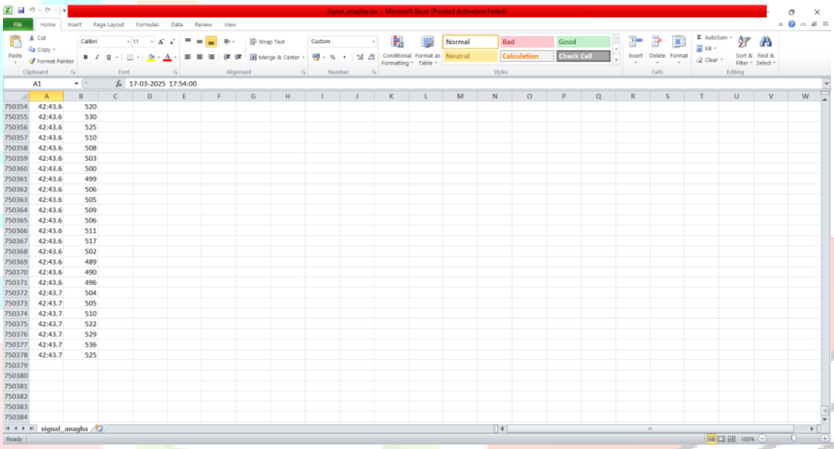

Fig.8 The above figure shows the data collected from the brain and is stored in the csv format.

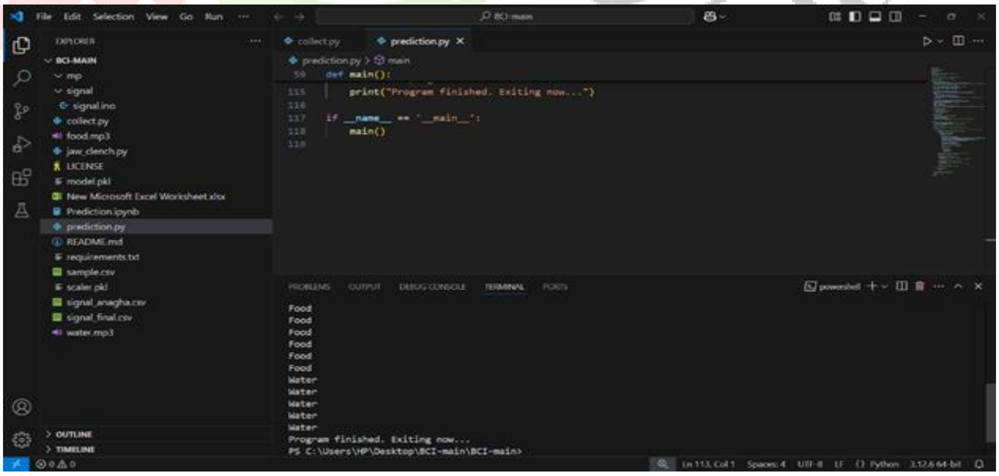

Fig.9 Output displayed in the terminal of VS Code

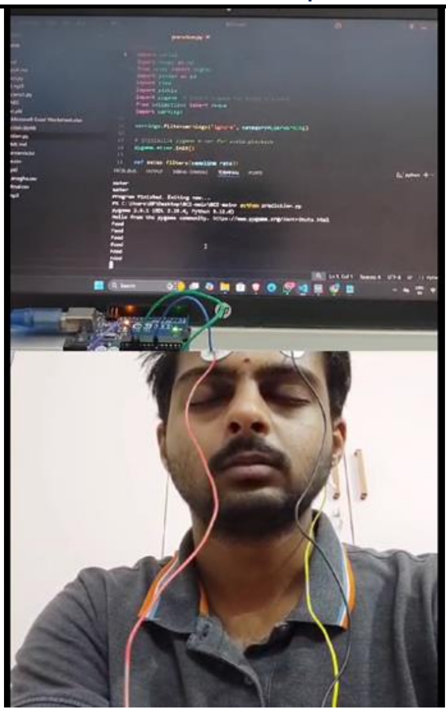

Fig.10 The system is detecting EEG signals and continuously predicting the word "food" based on the user's brain activity.

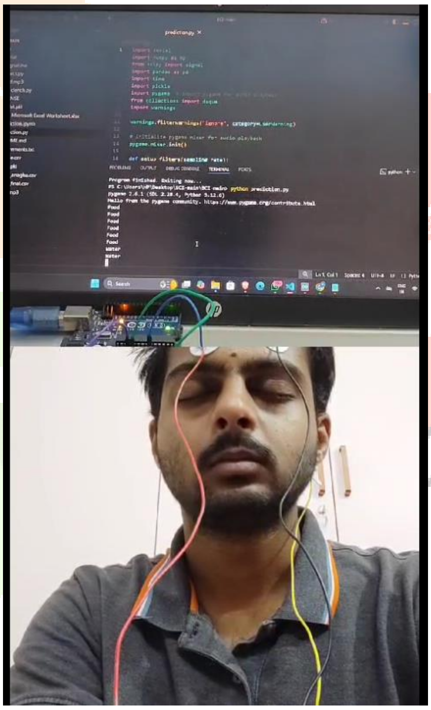

Fig.11 The system initially predicts "food" multiple times and then successfully identifies "water," showcasing accurate classification of different thoughts.

VII. CONCLUSION

This project successfully demonstrates a Brain-Computer Interface (BCI) system that enables individuals with severe speech impairments to communicate using EEG signals. By integrating EEG electrodes, Bio Amp EXG Pill, Arduino Uno, and deep learning models, the system accurately translates brain activity into speech. Real-time signal processing through Google Colab and Visual Studio Code ensures precise classification and output generation. Rigorous testing and optimization enhance accuracy and adaptability. Future improvements, including wearable EEG headsets and multilingual capabilities, will further enhance usability. This innovation paves the way for advanced assistive communication technologies, improving the quality of life for affected individuals.

Additionally, a timed audio playback mechanism has been seamlessly incorporated, ensuring that sounds play at predefined intervals. This modification blends naturally into the existing EEG-based classification framework, making the system robust and presentation-ready.

The project demonstrates the potential of real-time EEG signal analysis for practical applications, including human-computer interaction, neurofeedback training, and assistive technologies. It highlights the integration of AI and bio signals for intelligent decision-making, paving the way for future advancements in brain-computer interfaces.

VIII. FUTURE SCOPE

- Developing IOT applications where simple mobility movements can be done like switching on fan, sounding an alarm or switching on/off light.
- Creating emotional aware communication- Recognizing major emotions like sadness anger and fear so that caretakers can tend to the needs of individuals.
- Creation of open source or low-cost educational kit to teach students about the integration of EEG and AI.
- Partnering up with hospital and neurotherapy centres for real world application and trials of the project.

REFERENCES

[1] J. Lee, A. Kommineni, T. Feng, K. Avramidis, X. Shi, S. R. Kadiri, and S. Narayanan, “Toward Fully-End-to-End Listened Speech Decoding from EEG Signals,” *arXiv preprint*, arXiv:2406.08644, Jun. 2024.
[2] Z. Zhang, X. Ding, Y. Bao, Y. Zhao, X. Liang, B. Qin, and T. Liu, “Chisco: An EEG-Based BCI Dataset for Decoding of Imagined Speech,” *2024*.
[3] N. S. Amer and S. B. Belhaouari, “EEG Signal Processing for Medical Diagnosis, Healthcare, and Monitoring: A Comprehensive Review,” *Dec. 21, 2023*.
[4] M. A. Pfeffer, S. S. H. Ling, and J. K. W. Wong, “Exploring the Frontier: Transformer-Based Models in EEG Signal Analysis for Brain-Computer Interfaces,” *2024*.
[5] R. Kannan, A. Thulasi, Rithun, M. K. V. Manoj, A. G. B. V., and U. Dutt, “Thought Controlled Wheelchair Using EEG Acquisition Device,” *2023*.
[6] “Continuous and Discrete Decoding of Overt Speech with Non-Invasive EEG,” *bioRxiv preprint*, *2024*.
[7] S. Kim, Y.-E. Lee, S.-H. Lee, and S.-W. Lee, “Diff-E: Diffusion-Based Learning for Decoding Imagined Speech EEG,” *arXiv preprint*, arXiv:2307.14389, Jul. 2023.
[8] Y.-E. Lee, S.-H. Lee, S.-H. Kim, and S.-W. Lee, “NeuroTalk: Towards Voice Reconstruction from EEG During Imagined Speech,” *arXiv preprint*, arXiv:2301.07173, Jan. 2023.