International Journal of Scientific Research in Science and Technology
Available online at : www.ijsrst.com
Print ISSN: 2395-6011 | Online ISSN: 2395-602X
doi : https://doi.org/10.32628/IJSRST2411581

Forensic Perspective on Voice Biometrics and AI : A Review

Pragati Jain[1], Poorvi Ujjainia[1], Anshika Srivastava[2], Kajal Shrivastav[1], Ishu Rani[3], Akshat Vashisht[4], Rudranarayan Behera[1], Bhavika Moza[1], Debhjit Mukherjee[3*]

[1]Department of forensic science, University Institute of applied health sciences, Chandigarh university, NH-95, Ludhiana, Punjab, India

[2]Department of Forensic Science, College of Paramedical Sciences, Teerthanker Mahaveer University, Moradabad, Uttar Pradesh, India

[3]Department of Forensic Sciences, Chandigarh School of Business, Chandigarh Group of Colleges, Jhanjeri, Punjab, India

[4]Department of Computer application, Silver Oak College of Computer Application (SOCCA), Silver Oak university, Ahmedabad, Gujarat, India

***Corresponding Author Contact Details:** debhjit.mukherjee2018@gmail.com

ARTICLEINFO

Article History:

Accepted : 20 Aug 2024

Published : 04 Sep 2024

Publication Issue :

Volume 11, Issue 5

Sep-Oct-2024

Page Number :

49-63

ABSTRACT

Modern internet has given rise to various voice related crimes worldwide, notably deepfake voice scams where the perpetrators utilize artificial intelligence to deceive victims by the means of forgery of voice. This review article aims to discuss the advancements and challenges in voice biometrics, particularly focusing on the impact of AI and deep learning on this field. It underscores the evolution of voice biometrics from early methods to modern AI enhanced techniques, by highlighting the significant improvements in accuracy, security, and adaptability etc. The key findings of the article have highlighted that while AI-driven advancements have addressed many challenges including voice robustness and multilingual recognition, new threats like deep fake audio require ongoing innovation. The integration of various methods like deep learning, neural networks and advanced feature extraction has shown incredible potential in enhancing the system resilience. But challenges such as voice variability, privacy concerns and the forensic applications of these technologies remain critical issue to be addressed by the future researchers. This review article recommends multidisciplinary research to bridge the gap between this field and forensic science, emphasizing the need for continued development to address and prevent emerging threats very efficiently.

Keywords: Voice Biometrics, Artificial Intelligence (AI), Deepfake Voice Scams, Deep Learning, Neural Networks, Multilingual Recognition, Forensic Applications

I. INTRODUCTION

Easily accessible internet, deep learning methods have given rise to voice related crimes all over the world. Different forgery methods created by artificial intelligence approaches, poses a great threat, especially to speaker verification systems [1]. The preparators are utilizing artificial intelligence to produce a voice of victim's family members and on account of that, victims are falling in these types of frauds [2]. This type of phenomena is known as Deepfake voice scams. It is defined as the fake voices or spoofing used by perpetrators to deceive others by conveying false information. Firstly, Voice synthesis is the key element which is used to produce fake voices. It enables the perpetrator in transforming the attacker's voice to match the targets. In the other hand, the second method utilized by the criminals are Text-to-speech method for spoofing. Due to this method the identification of the perpetrator becomes very much harder, all because of the input text in the place of voice [3]. Third, AI tools like Voice-lab are being utilized to process person's voice biometrics, resulting in a deepfake voice crime, it is known as voice cloning. It can be generated by using very small amounts of biometric information harvested from personal content published in public sources such as TikTok, Facebook, Instagram, and other platforms, including government portals. Due to these types of crimes in June, 2023, FBI warned the public about the cyber criminals using the deepfake technologies to manipulate benign photos and videos in sextortion schemes, which have been lucrative for cybercriminals. According to the Federal Trade Commission, impostor scams accounted for the second highest reported losses in 2022, which amounted to US$2.6 billion [4]. In 2023. About 83% Indians have lost money in AI voice scams. In India, according to the reports of National Crime Records bureau (NCRB), the national capital Delhi registered 685 cybercrime cases in 2022 as against 345 in 2021 and 166 in 2020 related to voice cloning and AI [5]. According the survey report of McAfee survey, majority of the Indians have been found very susceptible to these types of scams. In this report it has been found that 66% of the respondents in India reacts to phone calls related to urgent financial help, from familiar caller to be relative like their parents (46%), spouse (34%), child (12%). The survey has also highlighted the most used convincible pretext utilized the scammers like got robbed used by 70% of the scammers, involved in a car crash used by (69%) of the scammers, loss of phone or wallet used by (65%) of the scammers, needing fund during international travel used 62% percent of the scammers. This report also signified that about 86% of the Indian population shares their voice data online or through the medium of voice messages in atleast once a week [6]. To prevent these types of voice-based crimes, advanced voice biometric systems are being employed in various health care sectors, remote-banking sectors, government sectors, transport sectors etc [7].

II. EVOLUTION OF VOICE BIOMETRICS

The evolution of biometric systems, particularly in voice recognition, has seen significant advancements over the years, marked by a focus on improving accuracy, security, and adaptability to emerging challenges. Revett (2008) laid the groundwork by introducing behavioral biometrics, highlighting its potential as an alternative to traditional identification methods [8]. Building on this foundation, Atah and Howells (2008) proposed template-free voice-based encryption to enhance security without storing sensitive data [9]. Urias et al. (2007) explored modular neural networks with type-2 fuzzy logic, improving recognition performance, while Chetty (2010) advanced the field with fuzzy fusion methods for liveness detection [9], [10]. Fong (2012) contributed computational algorithms for voice classification, employing techniques like hierarchical clustering and decision trees, setting the stage for more complex biometric applications [11]. Zhu et al. (2013) addressed security concerns with anti-spoofing

methods, critical for preventing fraudulent attacks, which were further enhanced by Luckyanets et al. (2017) with the development of robust liveness detection techniques [12], [13]. The exploration of unique biometric identifiers continued with Sailor et al. (2015), who introduced humming as a novel biometric method, leveraging the Modified Group Delay Function (MODGDF) for improved recognition accuracy [14]. This was complemented by Garlapati et al. (2016), who tackled tamper detection in speech-based access control systems through Log Coordinate Mapping (LCM) to combat DA-AD (Digital to analog-analog to digital) conversion attacks [15].

As voice biometrics gained traction, the reliability and security of these systems became a focal point. Van Rensburg and Von Solms (2017) evaluated voice recognition as a user-authentication method, suggesting the integration of AI for enhanced reliability [16]. Simultaneously, Soviany et al. (2017) explored multimodal biometrics, combining fingerprint, iris, and voice recognition to improve security while reducing computational demands [17]. Arslan et al. (2017) provided a comprehensive review of machine learning techniques, reflecting the broader integration of AI into biometric systems [18]. The critical issue of replay spoofing attacks in voice biometrics was addressed by Mankad et al. (2019), who improved feature selection methods to reduce error rates [19]. The growing threat of AI-driven cyberattacks, particularly deepfakes, was recognized by Ring (2021), with Bernal-Romero et al. (2024) proposing Physical Unclonable Functions (PUFs) as a countermeasure [20], [21]. This aligns with earlier work by AbdAlmageed (2019), who emphasized the importance of anti-spoofing measures in modern voice authentication systems [22].

As the COVID-19 pandemic underscored the need for trust in contactless interactions, Kathuria et al. (2020) explored factors influencing trust in voice biometrics, advocating for multi-level authentication and vernacular voices to boost user confidence [23]. In parallel, Evsyukov et al. (2021) offered an updated classification of anti-spoofing measures, identifying key research areas for further advancement [24]. Shakil et al. (2021) and Chato & Latifi (2018) contributed to the enhancement of system performance through the application of machine learning and deep learning algorithms, particularly in gender classification and large dataset analysis [25], [26]. Singh (2019) and García-Mateo & Chollet (2021) expanded on Revett's (2008) work by integrating speech technology into man-machine interfaces and large-scale applications, respectively, demonstrating the growing utility of biometrics across various domains [7], [8], [27].

In more recent years, Srinivasan (2023) critically examined the ethical and privacy concerns surrounding biometric advancements, highlighting the importance of legal safeguards and public awareness [28]. Ulutas et al. (2023) addressed the emerging threat of deepfake audio, emphasizing the need for advanced detection methods in speaker verification [1]. Aanjanadevi et al. (2023) proposed a system that combines fuzzy extractors and convolutional neural networks to further support biometric security and performance [29]. The integration of biometrics into consumer technology was exemplified by Harilatha and Riyazuddin (2020), who applied biometric authentication to home security systems using Amazon Alexa. Kumar (2022) reinforced the shift toward advanced biometric solutions by discussing the role of biometrics and AI in enhancing business security [30].

The complexity of biometric performance in multilingual contexts, particularly in speaker identification, was highlighted by de Lima & Da-Costa Abreu (2022), who emphasized the challenges posed by linguistic diversity [31]. Meanwhile, Enomoto and Koito (2019) examined the future of multimodal biometrics with NEC Bio-IDiom, aligning with ongoing trends towards enhanced security and reliability [32].

Despite of the evolution in the field of biometrics and substantial advancements in voice biometrics,

particularly in enhancing accuracy, security, and integrating AI, a critical literature gap remains in the context of its application within forensic science. While numerous studies have focused on improving recognition performance, addressing security challenges, and integrating AI for enhanced reliability, there is limited research explicitly examining the forensic implications of these advancements. The current body of work largely overlooks the specific challenges and complexities encountered when applying voice biometrics in legal contexts, such as the admissibility of voice evidence, the effectiveness of anti-spoofing measures in forensic investigations, and the ethical considerations unique to the forensic domain. Additionally, there is a need for more interdisciplinary research that bridges the gap between AI advancements and forensic applications, particularly in addressing emerging threats like deepfake audio and ensuring robust, legally sound methodologies for voice evidence. This review aims to fill this gap by critically analyzing these overlooked aspects, those integrate technological innovations with forensic requirements.

III. FUNDAMENTAL OF VOICE BIOMETRICS

Personal identification necessitates the recognizing of the individuals based on their physical attributes [33]. Biometrics delivers a reliable and systematic approach to this by measuring and analyzing the biological data, with voice biometrics being significantly valuable due to its individualistic physiological and behavioral traits shaped by the mouth and throat [34], [35]. The production of the voice occurs through the rapid vibration of vocal cords within the larynx, generating sound waves [36]. The pitch, determined by vocal cord tension, varies from low frequencies in male voices to high frequencies in female soprano voices, and beyond in trained singers. The harmonic structure, formed by overtones, defines the timbre, while formants, influenced by the shape of the throat and nasal passages, are crucial for vowel perception. Voices are categorized into six types—bass, baritone, tenor for men, and contralto, mezzo-soprano, soprano for women—based on factors such as gender, body type, and age, which determine vocal range and tonal quality. Forensic analysis compares voice samples to categorize them as belonging to either the same speaker or different speakers, aiding in the identification process [37], [38]

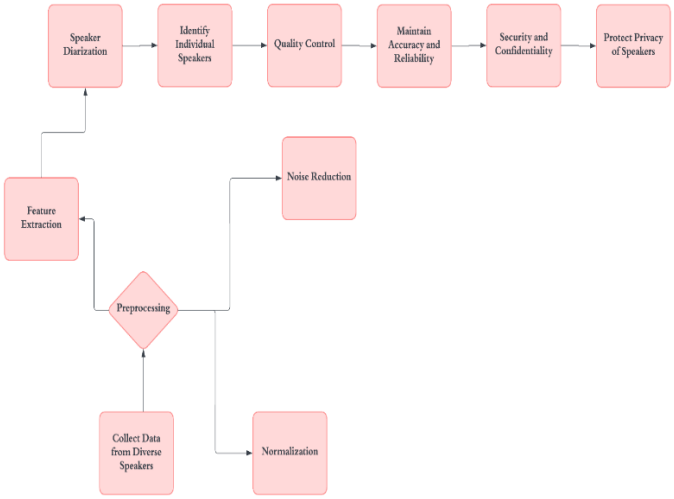

Figure 1: The figure represents the overview of the process of the speaker identification. (Source: Author)

IV. ROLE OF AI IN VOICE BIOMETRICS

In the 1968 film 2001: A Space Odyssey, HAL 9000 analyzes speech to detect mood, a concept that seemed futuristic at the time. Now, with advances in AI and sensing technologies, such capabilities are no longer fiction but reality [22].

Feature extraction is crucial for the accuracy and performance of speaker identification (SI) and biometric recognition systems. Significant studies have highlighted the dominance of Mel-Frequency Cepstral Coefficients (MFCCs) in SI, noting a trend toward robust frameworks addressing challenges like noise robustness and multi-lingual recognition [38]. Addressing security concerns, it has been found that entropy-based feature selection significantly improved SI systems' resistance to replay attacks [19]. Several research studies have signified on-hand vein recognition and SI under real-world conditions

further supporting the central role of feature extraction in biometric systems [39], [40].

Integration of machine learning in voice biometric recognition technique focuses on effective feature extraction, which significantly impacts recognition performance. For example, [42] compare Linear Predictive Coefficients (LPC) and Prosodic Features for voice recognition. Machine learning models such as Radial Basis Function Network (RBFN), Support Vector Machines (SVM), Artificial Neural Networks (ANN), and Transfer Learning Convolutional Neural Networks (TL-CNN) are commonly employed for accurate biometric identification [42], [43], [44].

Another significant trend is the adoption of deep learning techniques across these studies, especially for tasks involving image and signal processing. Additionally, [45] explore machine learning cluster analysis in intrusion detection systems, illustrating how deep learning can enhance biometric security. Furthermore, studies like those by [46] and [47]explore the integration of advanced sensors and machine learning for improved accuracy in biometric systems, emphasizing the growing role of these technologies in developing robust, reliable identification systems for forensic and security applications.

A. Machine Learning and Deep Learning in Voice Biometric Recognition

Recent advancements in artificial intelligence and machine learning have significantly enhanced biometric systems, as demonstrated in various studies. [48] discusses the potential of machine learning to improve the efficiency and reliability of biometric systems across physiological and behavioral traits. The COVID-19 pandemic has accelerated the adoption of touchless interfaces, with [49] exploring hand gesture interfaces and [50] focusing on Automatic Speaker Verification (ASV) systems for noisy environments. The growing threat of deep fake audio spoofing, which poses a significant risk to speaker verification systems, is addressed by [51] through a deep learning-based detection method. [52] explored voice biometrics by employing an innovative piezoelectric polymeric interface and neural network model, resulting in over 96% accuracy for population identification and speaker recognition and over 93% accuracy for healthcare assessment.

[53] addressed the challenge of spoofing attacks in biometric speech recognition by fusing physiological and physical features through SE-DenseNet, achieving a 5% improvement in the tandem decision cost function and a 7% enhancement in the equal error rate on the ASVspoof 2019 dataset. In another study, [54], [55] focused on cross-modal voice-to-face systems, introducing adversarial attacks that generated high target face-matching rates and demonstrated the potential vulnerabilities of these systems. [56] utilized transfer learning to enhance speaker verification for short-duration audio samples, achieving high accuracy in applications like voice assistants. [57] reviewed automatic speaker recognition (ASR) methods, highlighting the effectiveness of CNNs in improving accuracy for security monitoring and access control. These studies collectively highlight the advancements in biometric technologies, emphasizing multi-modal approaches, anti-spoofing measures, and the transition to contactless systems in response to emerging security challenges.

B. Support Vector Machines (SVM) in Enhancing Voice Biometric Security

Recent advancements in biometric systems have significantly improved authentication methods while addressing emerging challenges. Voice biometrics, though enhanced, still face vulnerabilities to artificial speech attacks. The Hybrid AIR-ASVspoof (HAIR-ASVspoof) approach, combining AIR-ASVspoof deep features with an SVM classifier, achieved a notable EER of 0.57%, enhancing robustness against spoofing [58]. This advancement is complemented by research on multilingual and multi-modal voice recognition, where [59] achieved 99.27% accuracy using MFCC and SVM, highlighting the effectiveness of advanced features in overcoming language-specific challenges.

Additionally, studies on formant characteristics for biometric authentication and emotion detection have demonstrated the versatility of voice biometrics beyond traditional uses [60], [61].

C. Neural Networks in Voice Biometrics

[62] optimized speaker identification (SI) by integrating a CNN-based system with an Enhanced Multi-Active Learner framework, surpassing benchmark systems with a 2.43% improvement in top-1 accuracy and reducing trainable parameters by up to 95%. Likewise, [63] tested the resilience of biometric identity verification systems against voice cloning attacks using advanced neural networks, finding that these systems could resist such attacks even when humans could not distinguish between original and cloned voices. [64] presented a speaker verification system combining CNN with STM32 Cube.AI, achieving 93.50% accuracy through dual authentication. [65] compared Siamese Neural Networks (SNN) with Gaussian Mixture Models (GMM), with SNNs achieving a lower Equal Error Rate of 4.88%. [66] improved Time Delay Neural Networks (TDNN) for speaker verification, enhancing performance in spoofing detection and verification tasks.

[67] proposed a single-speaker diarization system with over 96.4% accuracy in distinguishing multiple speakers.

[68] uses Gaussian Mixture Models (GMMs) and Universal Background Models (UBM) to enhance speaker recognition, while Convolutional Neural Networks (CNNs) learn speaker characteristics. By merging fingerprint minutiae with voice features, the system improves accuracy and reduces privacy risks.

[44] expands on this by implementing Transfer Learning Convolutional Neural Networks (TL-CNN) for multi-modal biometric security, incorporating multiple datasets such as retina, face, and voice. The system achieved superior performance in recognition and authentication through a bio-optimization-based modified Lion optimization algorithm (MLOA). [69] introduced multi-task neural networks (MTNs) for voice-based soft biometrics in social robots, achieving state-of-the-art performance in gender, age, and emotion recognition tasks, optimized for low-power devices.

[50] addressed the need for robust Automatic Speaker Verification (ASV) systems during the COVID-19 pandemic by improving spoken language identification under noisy and spoofed conditions. They combined Constant Q cepstral coefficients- Mel-Frequency Cepstral Coefficients (CQCC-MFCC) features with a CNN to enhance performance in multilingual environments, achieving 97% accuracy on the INDIC TTS Database [50].

D. LSTM's Role in Countering Spoofing Attacks

Automatic speaker verification (ASV) systems are prone to spoofing attacks like voice conversion and replayed speech. LSTM networks have significantly advanced the field of voice biometrics, particularly in enhancing the robustness of Automatic Speaker Verification (ASV) systems against spoofing attacks. [70] study introduces regional energy features, which analyze non-overlapping time-frequency regions, outperforming traditional methods. Tested on ASVspoof 2019 and 2021 datasets using models like LCNN-LSTM and SE-Res2Net, these features significantly improved spoof detection with minimal computational impact, proving effective in capturing key nonspeech and band-limited information for identifying spoofed speech. [71] developed a countermeasure using Gammatone Cepstral Coefficients (GTCC) and Mel-Frequency Cepstral Coefficients (MFCC), coupled with a Bi-directional LSTM classifier, achieving 97% accuracy and a low Equal Error Rate (EER) of 2.95%, demonstrating LSTM's effectiveness in distinguishing between genuine and spoofed speech. [72] similarly focused on creating a robust ASV system using LSTM for classification, showing impressive results with EERs of 2.38% and 2.72%, and accuracy rates of 98% and 97.26% under different attack scenarios. LSTM's

application extends beyond ASV, as shown by [73], who achieved a remarkable 99.98% accuracy in recognizing individuals based on Bangla speech, highlighting LSTM's capability to handle complex, language-specific data. However, [74] found that Dense Neural Networks outperformed LSTM in speaker prediction tasks, with the former achieving over 95% accuracy compared to LSTM's 79%, suggesting that while LSTM is effective, other architectures might be better suited for specific applications. Hybrid approaches, such as the one proposed by [75], combining CNN and LSTM, have also shown promise in enhancing the detection of multi-replay attacks, with promising results in terms of accuracy and low EER. These studies collectively underscore LSTM's critical role in voice biometrics while also highlighting the potential benefits of integrating or comparing LSTM with other neural network architectures for optimal performance.

E. Advancements in Recurrent Neural Networks (RNNs) and LSTM for Voice Biometrics

Recurrent Neural Networks (RNNs) have significantly advanced the field of voice biometrics, providing solutions to challenges such as environmental noise and enhancing multimodal biometric systems. [76] demonstrated the effectiveness of a two-layer bi-directional Long Short-Term Memory (LSTM) RNN model in detecting unusable responses caused by background noise in spoken language proficiency tests, achieving superior performance compared to traditional models like Gaussian Mixture Models (GMM) and AdaBoost classifiers. The utility of RNNs in processing and integrating multiple biometric signals was further evidenced by [77], who developed a multimodal system combining facial and voice recognition using LSTM networks, resulting in a significantly higher accuracy than unimodal systems. However, as the complexity of voice biometrics increased, limitations of RNNs, particularly in capturing long-range dependencies, became apparent. This led to the exploration of transformer-based models by [78], which improved speaker recognition performance across varied accents through advanced data augmentation and denoising techniques. While transformers represent an evolution in voice biometrics, the foundational contributions of RNNs, especially in enhancing robustness in noisy environments and multimodal contexts, remain crucial in the ongoing development of secure and reliable biometric systems.

Liveness detection is vital in biometrics, particularly for preventing spoofing in voice recognition systems like Automatic Speaker Verification (ASV). Traditional focus has been on face and iris recognition, with less attention to Voice Liveness Detection (VLD). VLD can identify spoofing by detecting "pop noise," a subtle sound from live speakers, as a sign of authenticity. [79] explored the Constant-Q Transform (CQT) for VLD, which excels in detecting low-frequency sounds like pop noise, outperforming traditional methods like the Short-Time Fourier Transform (STFT). On the POCO dataset, CQT-based methods, when combined with classifiers like SVM, GMM, and various deep learning models like CNN, Light-CNN (LCNN), and Residual Networks (ResNet), significantly improved accuracy, showing up to 14.23% better performance than STFT.

[51] also used CQT to detect deep fake audio, generating spectrogram images that were classified with a vision transformer network. Their approach, tested on the ASVSpoof 2019 dataset, effectively identified fake audio, further proving CQT's effectiveness in voice biometrics. Together, these studies demonstrate CQT's strong potential in improving both liveness detection and deep fake detection in voice recognition systems.

TABLE 1 CHALLENGES AND LIMITATIONS OF AI IN VOICE BIOMETRICS

Author Profile	Challenges	Limitations	Reference
Evsyukov et al. 2021	Voice biometric systems are vulnerable to spoofing attacks using recorded or synthetic voices.	Current antispoofing techniques may not effectively address sophisticated attacks, with effectiveness varying by type.	[80], [81]
Babu et al. 2024	Variability in voice due to accent, speech disorders, and emotional states affects system accuracy.	Systems may struggle to maintain high performance across diverse populations and changing voice conditions.	[82]
Parlar 2023	The collection and storage of voice data raise significant privacy concerns, as voiceprints are sensitive. AI models must generalize to new and unseen data but often struggle with new types of attacks or voice pattern changes.	Ensuring the security of voice data and protecting it from unauthorized access or misuse remains a critical issue. Performance may degrade when exposed to novel spoofing techniques or diverse acoustic environments.	[83]

V. COUNTERMEASURES

To enhance the security and effectiveness of voice biometric systems, several key strategies should be implemented. Advanced antispoofing techniques, such as deep learning-based detection and multimodal approaches, offer improved resistance to spoofing attacks by integrating methods like replay attack detection and feature extraction improvements [41]. In parallel, robust data protection protocols are crucial; employing strong encryption and secure storage practices can significantly reduce risks associated with data breaches, while regular updates and audits ensure compliance with privacy laws [44]. Additionally, maintaining the effectiveness of voice biometric systems requires regular model updates and extensive testing on emerging attack vectors. Incorporating feedback mechanisms and adaptive learning techniques helps refine models and improve performance over time [45].

VI. CONCLUSION

The field of voice biometrics has witnessed various advancements, driven by several innovations in the field of artificial intelligence and machine learning, which has increased the reliability, security, and adaptability of this field. The evolution from traditional voice recognition methods to new AI integrated techniques have addressed many challenges such as noise robustness, multi-lingual recognition, and resistance to spoofing attacks. But,

the recent emergence of deepfake voice scams and sophisticated spoofing methods signifies a need for continuous innovations to develop more robust anti-spoofing methods and improving the resilience of the voice biometric system. The integration of deep leaning, neural networks and advanced feature extraction techniques will show a great future perspective to future generation forensic experts. Yet, challenges such as variability in the voice due to accents, emotional sates and ethical implications of data privacy remain some critical issues to be addressed. The literature gap in the forensic application of voice biometrics integrated with AI reveals a serious need for multidisciplinary researches, those will bridge technological advancements with legal and ethical considerations. Acknowledging and addressing these gaps involves not only improving the technical aspect but also signifying the forensic applications meet the legal standards and effectively counteract the emerging threats worldwide.

VII. REFERENCES

[1]. G. Ulutas, G. Tahaoglu, and B. Ustubioglu, "Deepfake audio detection with vision transformer based method," in 2023 46th International Conference on Telecommunications and Signal Processing, TSP 2023, 2023, pp. 244–247. doi: 10.1109/TSP59544.2023.10197715.

[2]. Times of India, "About 83% Indians have lost money in AI voice scams: Report ," 2023.

[3]. Y. Yanagi, R. Orihara, Y. Tahara, Y. Sei, T. Alumäe, and A. Ohsuga, "The Proposal of Countermeasures for DeepFake Voices on Social Media Considering Waveform and Text Embedding," Annals of Emerging Technologies in Computing, vol. 8, no. 2, pp. 15–31, 2024, doi: 10.33166/AETiC.2024.02.002.

[4]. Craig Gibson and Josiah Hagen, "Virtual Kidnapping ," https://www.trendmicro.com/vinfo/us/security/news/cybercrime-and-digital-threats/how-cybercriminals-can-perform-virtual-kidnapping-scams-using-ai-voice-cloning-tools-and-chatgpt.

[5]. NDTV, "AI Voice Cloning: What It Is And How To Avoid Getting Scammed By It," Feb. 2024.

[6]. M. Guest Author, "Voice Cloning Scams: The Alarming Use of Artificial Intelligence by Cybercriminals," https://www.medianama.com/2024/04/223-voice-cloning-scams-cybercriminals-ai/.

[7]. C. García-Mateo and G. Chollet, Voice Biometrics: Technology, trust and security. 2021. doi: 10.1049/PBSE012E.

[8]. K. Revett, Behavioral Biometrics: A Remote Access Approach. 2008. doi: 10.1002/9780470997949.

[9]. J. A. Atah and G. Howells, "Score normalisation of voice features for template free biometric encryption," in International Conference on Artificial Intelligence and Pattern Recognition 2008, AIPR 2008, 2008, pp. 269–272. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84876774976&partnerID=40&md5=51cd3d5a316bcd1aa9b0b5b3a3ced5e7

[10]. G. Chetty, "Biometric liveness checking using multimodal fuzzy fusion," in 2010 IEEE World Congress on Computational Intelligence, WCCI 2010, 2010. doi: 10.1109/FUZZY.2010.5584864.

[11]. S. Fong, "Using hierarchical time series clustering algorithm and wavelet classifier for biometric voice classification," J Biomed Biotechnol, vol. 2012, 2012, doi: 10.1155/2012/215019.

[12]. Z. Y. Zhu, Q. H. He, X. H. Feng, Y. X. Li, and Z. F. Wang, "Liveness detection using time drift between lip movement and voice," in Proceedings - International Conference on Machine Learning and Cybernetics, 2013, pp. 973–978. doi: 10.1109/ICMLC.2013.6890423.

[13]. E. Luckyanets, A. Melnikov, O. Kudashev, S. Novoselov, and G. Lavrentyeva, "Bimodal anti-spoofing system for mobile security," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2017, pp. 211–220. doi: 10.1007/978-3-319-66429-3_20.

[14]. H. B. Sailor, M. C. Madhavi, and H. A. Patil, "Significance of phase-based features for person recognition using humming," in ACM International Conference Proceeding Series, 2015, pp. 99–103. doi: 10.1145/2708463.2709035.

[15]. B. M. Garlapati, S. R. Chalamala, and K. R. Kakkirala, "Tamper detection in speech based access control systems using watermarking," in Proceedings - AIMS 2015, 3rd International Conference on Artificial Intelligence, Modelling and Simulation, 2016, pp. 325–331. doi: 10.1109/AIMS.2015.59.

[16]. E. O. J. Van Rensburg and R. Von Solms, "Voice recognition as a user-authentication method," in European Conference on Information Warfare and Security, ECCWS, 2017, pp. 702–709. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85028025006&partnerID=40&md5=4fcfa6cc4f6a6b6fadb1b8567ff8f705

[17]. S. Soviany, V. Săndulescu, and S. Puşcoci, "A multimodal biometric identification method for mobile applications security," in Proceedings of the 8th International Conference on Electronics, Computers and Artificial Intelligence, ECAI 2016, 2017. doi: 10.1109/ECAI.2016.7861102.

[18]. B. Arslan, E. Yorulmaz, B. Akca, and S. Sagiroglu, "Security perspective of Biometric recognition and machine learning techniques," in Proceedings - 2016 15th IEEE International Conference on Machine Learning and Applications, ICMLA 2016, 2017, pp. 492–497. doi: 10.1109/ICMLA.2016.183.

[19]. S. H. Mankad, S. Garg, M. Patel, and H. Adalja, "Investigating Feature Reduction Strategies for Replay Antispoofing in Voice Biometrics," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2019, pp. 400–408. doi: 10.1007/978-3-030-34872-4_44.

[20]. T. Ring, "Europol: the AI hacker threat to biometrics," Biometric Technology Today, vol. 2021, no. 2, pp. 9–11, 2021, doi: 10.1016/S0969-4765(21)00023-0.

[21]. J. C. Bernal-Romero, J. M. Ramirez-Cortes, and J. De Jesus Rangel-Magdaleno, "Unbreakable Biometrics: How Physical Unclonable Functions are Revolutionizing Security," IEEE Instrum Meas Mag, vol. 27, no. 2, pp. 71–78, 2024, doi: 10.1109/MIM.2024.10472986.

[22]. W. AbdAlmageed, "Anti-spoofing's state-of-the-art: deep learning meets super sensors," Biometric Technology Today, vol. 2019, no. 7, pp. 8–11, 2019, doi: 10.1016/S0969-4765(19)30099-2.

[23]. R. Kathuria, A. Wadehra, and V. Kathuria, "Human-Centered Artificial Intelligence: Antecedents of Trust for the Usage of Voice Biometrics for Driving Contactless Interactions," in Communications in Computer and Information Science, 2020, pp. 325–334. doi: 10.1007/978-3-030-60700-5_42.

[24]. M. V Evsyukov, M. M. Putyato, and A. S. Makaryan, "Antispoofing Countermeasures in Modern Voice Authentication Systems," in CEUR Workshop Proceedings, 2021, pp. 197–202. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85122794382&partnerID=40&md5=6e038815dd07d23968182b493ff3a593

[25]. S. Shakil, D. Arora, and T. Zaidi, "Feature based classification of voice based biometric data

through Machine learning algorithm," in Materials Today: Proceedings, 2021, pp. 240–247. doi: 10.1016/j.matpr.2021.05.261.

[26]. L. Chato and S. Latifi, "Application of Machine Learning to Biometric Systems- A Survey," in Journal of Physics: Conference Series, 2018. doi: 10.1088/1742-6596/1098/1/012017.

[27]. S. Singh, "The role of speech technology in biometrics, forensics and man-machine interface," International Journal of Electrical and Computer Engineering, vol. 9, no. 1, pp. 281–288, 2019, doi: 10.11591/ijece.v9i1.pp281-288.

[28]. S. Srinivasan, "Understanding User Perception of Biometric Privacy in the Era of Generative AI," in 4th International Conference on Communication, Computing and Industry 6.0, C216 2023, 2023. doi: 10.1109/C2I659362.2023.10430931.

[29]. S. Aanjanadevi, S. Aanjankumar, K. R. Ramela, and V. Palanisamy, "Face Attribute Convolutional Neural Network System for Data Security with Improved Crypto Biometrics," Computer Systems Science and Engineering, vol. 45, no. 3, pp. 2351–2362, 2023, doi: 10.32604/csse.2023.031893.

[30]. G. Harilatha and K. Riyazuddin, "Artificial intelligence oriented security system using alexa," in Lecture Notes in Electrical Engineering, vol. 643, 2020, pp. 303–312. doi: 10.1007/978-981-15-3125-5_32.

[31]. T. A. de Lima and M. C. Da-Costa Abreu, "Phoneme analysis for multiple languages with fuzzy-based speaker identification," IET Biom, vol. 11, no. 6, pp. 614–624, 2022, doi: 10.1049/bme2.12078.

[32]. M. Enomoto and T. Koito, "Bio-idiom - NEC's biometric authentication brand," NEC Technical Journal, vol. 13, no. 2, pp. 14–18, 2019, [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85078325693&partnerID=40&md5=32a2b097c154bb438570bb0f8fd50fe6

[33]. F. Sal, Biometric Techniques For Personal Identification & Voice Authentication. 2019. doi: 10.13140/RG.2.2.16351.61607.

[34]. A. K. Jain and A. Kumar, "Biometrics of next generation: An overview," Second generation biometrics, vol. 12, no. 1, pp. 2–3, 2010.

[35]. A. K. Jain, A. Ross, and S. Pankanti, "Biometrics: a tool for information security," IEEE transactions on information forensics and security, vol. 1, no. 2, pp. 125–143, 2006.

[36]. N. Singh, A. Agrawal, and R. A. Khan, "Voice biometric: A technology for voice based authentication," Adv Sci Eng Med, vol. 10, no. 7–8, pp. 754–759, 2018.

[37]. P. Jain, P. Chinmayee, K. Kaur, S. Chaudhary, K. Kaur, and S. Karunya, "Advancements in Forensic Voice Analysis: Legal Frameworks and Technology Integration," Asian Journal of Advances in Research, vol. 7, no. 1, pp. 369–384, 2024.

[38]. S. Srivastava, A. A. Hussain, and S. Gupta, "A Review Article on Layered Voice Analysis: Forensic Utility and Limitation," International Journal of Indian Psychology, vol. 10, no. 3, 2022.

[39]. S. S. Tirumala, S. R. Shahamiri, A. S. Garhwal, and R. Wang, "Speaker identification features extraction methods: A systematic review," Expert Syst Appl, vol. 90, pp. 250–271, 2017, doi: 10.1016/j.eswa.2017.08.015.

[40]. A. Merouane, S. Benziane, P. Boulet, A. El Hassan Benyamina, and L. Loukil, "Hybridization of discrete binary particle swarm optimization and invariant moments for dorsal hand vein feature selection," in 2013 International Conference on Electronics, Computers and Artificial Intelligence, ECAI 2013, 2013. doi: 10.1109/ECAI.2013.6636192.

[41]. E. L. Campbell, G. Hernández, and J. R. Calvo, "Feature extraction of automatic speaker

recognition, analysis and evaluation in real environment," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2018, pp. 376–383. doi: 10.1007/978-3-030-01132-1_43.

[42]. V. S. Baidwan and S. Gujral, "Comparative analysis of prosodic features and linear predictive coefficients for speaker recognition using machine learning technique," in 2014 International Conference on Devices, Circuits and Communications, ICDCCom 2014 - Proceedings, 2014. doi: 10.1109/ICDCCom.2014.7024705.

[43]. K. K. Patro and P. R. Kumar, "Machine learning classification approaches for biometric recognition system using ECG signals," Journal of Engineering Science and Technology Review, vol. 10, no. 6, pp. 1–8, 2017, doi: 10.25103/jestr.106.01.

[44]. A. Gona, M. Subramoniam, and R. Swarnalatha, "Transfer learning convolutional neural network with modified Lion optimization for multimodal biometric system," Computers and Electrical Engineering, vol. 108, 2023, doi: 10.1016/j.compeleceng.2023.108664.

[45]. C. Turner and A. Joseph, "A Statistical and Cluster Analysis Exploratory Study of Snort Rules," in Procedia Computer Science, 2017, pp. 106–115. doi: 10.1016/j.procs.2017.09.023.

[46]. S. M. H. Mousavi and A. Ilanloo, "Seven Staged Identity Recognition System Using Kinect V.2 Sensor," in 2022 9th Iranian Joint Congress on Fuzzy and Intelligent Systems, CFIS 2022, 2022. doi: 10.1109/CFIS54774.2022.9756435.

[47]. B. Arslan, E. Yorulmaz, B. Akca, and S. Sagiroglu, "Security perspective of Biometric recognition and machine learning techniques," in Proceedings - 2016 15th IEEE International Conference on Machine Learning and Applications, ICMLA 2016, 2017, pp. 492–497. doi: 10.1109/ICMLA.2016.183.

[48]. L. Chato and S. Latifi, "Application of Machine Learning to Biometric Systems- A Survey," in Journal of Physics: Conference Series, 2018. doi: 10.1088/1742-6596/1098/1/012017.

[49]. I. El Magrouni, A. Ettaoufik, A. Siham, A. Maizate, and B. Lotfi, "Approach for the construction of gestural interfaces to control graphical interfaces based on artificial intelligence," in Proceedings - 2022 9th International Conference on Wireless Networks and Mobile Communications, WINCOM 2022, 2022. doi: 10.1109/WINCOM55661.2022.9966424.

[50]. A. R. Ambili and R. C. Roy, "Spoken Language Identification of Indian Languages in Adversarial Synthetic and Noisy Attacking Environments," in Proceedings of International Conference on Computing, Communication, Security and Intelligent Systems, IC3SIS 2022, 2022. doi: 10.1109/IC3SIS54991.2022.9885560.

[51]. G. Ulutas, G. Tahaoglu, and B. Ustubioglu, "Deepfake audio detection with vision transformer based method," in 2023 46th International Conference on Telecommunications and Signal Processing, TSP 2023, 2023, pp. 244–247. doi: 10.1109/TSP59544.2023.10197715.

[52]. A. Babu, E. Raoul, G. Kassahun, I. Dufour, D. Mandal, and D. Thuau, "Programmable Polymeric-Interface for Voiceprint Biometrics," Adv Mater Technol, vol. 9, no. 4, 2024, doi: 10.1002/admt.202301551.

[53]. J. Xue and H. Zhou, "Physiological-physical feature fusion for automatic voice spoofing detection," Front Comput Sci, vol. 17, no. 2, 2023, doi: 10.1007/s11704-022-2121-6.

[54]. Y. Chen, Y. Zhu, P. Zhao, and J. Guo, "Can you trust what you hear: Effects of audio-attacks on voice-to-face generation system," International Journal of Intelligent Systems, vol. 37, no. 5, pp. 3090–3116, 2022, doi: 10.1002/int.22825.

[55]. M. Chen, L. Lu, Z. Ba, and K. Ren, "PhoneyTalker: An Out-of-the-Box Toolkit for Adversarial Example Attack on Speaker Recognition," in Proceedings - IEEE INFOCOM, 2022, pp. 1419–1428. doi: 10.1109/INFOCOM48880.2022.9796934.

[56]. N. Fathima, J. B. Simha, and S. Abhi, "Transfer Learning for Speaker Verification with Short-Duration Audio," in Lecture Notes in Networks and Systems, 2024, pp. 195–205. doi: 10.1007/978-981-97-1313-4_17.

[57]. A. Dhole and V. Kadroli, "An Overview of Speaker Recognition: Conceptual Framework and CNN based Identification Technique," in 14th International Conference on Advances in Computing, Control, and Telecommunication Technologies, ACT 2023, 2023, pp. 2901–2908. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85174385196&partnerID=40&md5=d5dec00f0cfe1faae61a0e544fa2858e

[58]. C. B. Tan, M. H. Ahmad Hijazi, and P. N. Ellyza Nohuddin, "A Hybrid Classification Approach For Artificial Speech Detection," in 5th IEEE International Conference on Artificial Intelligence in Engineering and Technology, IICAIET 2023, 2023, pp. 236–240. doi: 10.1109/IICAIET59451.2023.10291764.

[59]. U. Sadique, M. S. Khan, S. Anwar, and M. Ahmad, "Machine Learning based human recognition via robust Features from audio signals," in 3rd IEEE International Conference on Artificial Intelligence, ICAI 2023, 2023, pp. 52–57. doi: 10.1109/ICAI58407.2023.10136683.

[60]. Y. Belova, "Comparative Analysis of Authentication Using Formant Features of Vowels and Consonants," in Springer Proceedings in Mathematics and Statistics, 2024, pp. 211–220. doi: 10.1007/978-3-031-52965-8_17.

[61]. S. Kumar et al., "Multilayer Neural Network Based Speech Emotion Recognition for Smart Assistance," Computers, Materials and Continua, vol. 74, no. 1, pp. 1523–1540, 2023, doi: 10.32604/cmc.2023.028631.

[62]. S. R. Shahamiri, "An optimized enhanced-multi learner approach towards speaker identification based on single-sound segments," Multimed Tools Appl, vol. 83, no. 8, pp. 24541–24562, 2024, doi: 10.1007/s11042-023-16507-2.

[63]. K. Milewski, S. Zaporowski, and A. Czyżewski, "Comparison of the Ability of Neural Network Model and Humans to Detect a Cloned Voice," Electronics (Switzerland), vol. 12, no. 21, 2023, doi: 10.3390/electronics12214458.

[64]. P. Wang et al., "Intelligent Access Control System Based on Voiceprint and Voice Technology," in Proceedings - 2022 11th International Conference of Information and Communication Technology, ICTech 2022, 2022, pp. 461–465. doi: 10.1109/ICTech55460.2022.00098.

[65]. P. Shetty, R. Rodricks, S. Malgundkar, H. Pamnani, and S. Katke, "SPEECH BIOMETRICS: A Comprehensive Deep Learning-based Speaker Identification System," in 3rd International Mobile, Intelligent, and Ubiquitous Computing Conference, MIUCC 2023, 2023, pp. 228–232. doi: 10.1109/MIUCC58832.2023.10278329.

[66]. J. Monteiro, J. Alam, and T. H. Falk, "Multi-level self-attentive TDNN: A general and efficient approach to summarize speech into discriminative utterance-level representations," Speech Commun, vol. 140, pp. 42–49, 2022, doi: 10.1016/j.specom.2022.03.008.

[67]. T. M. Al-Hadithy and M. Frikha, "A Real-Time Speaker Diarization System Based on Convolutional Neural Networks Architectures," in HORA 2023 - 2023 5th International Congress on Human-Computer Interaction, Optimization and Robotic Applications,

Proceedings, 2023. doi: 10.1109/HORA58378.2023.10156741.

[68]. A. Priyadharshini, R. Balakrishnan, S. Mohamed Shazuli, D. Gunapriya, and D. Joseph, "Convolutional Neural Network for Speaker Recognition Embedding with Biometric System," in 5th International Conference on Inventive Computation Technologies, ICICT 2022 - Proceedings, 2022, pp. 896–900. doi: 10.1109/ICICT54344.2022.9850483.

[69]. P. Foggia, A. Greco, A. Roberto, A. Saggese, and M. Vento, "Identity, Gender, Age, and Emotion Recognition from Speaker Voice with Multi-task Deep Networks for Cognitive Robotics," Cognit Comput, 2024, doi: 10.1007/s12559-023-10241-5.

[70]. G. Dişken, "Complementary regional energy features for spoofed speech detection," Comput Speech Lang, vol. 85, 2024, doi: 10.1016/j.csl.2023.101602.

[71]. J. Zhou, T. Hai, D. N. A. Jawawi, D. Wang, E. Ibeke, and C. Biamba, "Voice spoofing countermeasure for voice replay attacks using deep learning," Journal of Cloud Computing, vol. 11, no. 1, 2022, doi: 10.1186/s13677-022-00306-5.

[72]. S. Joshi and M. Dua, "LSTM-GTCC based Approach for Audio Spoof Detection," in 2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing, COM-IT-CON 2022, 2022, pp. 656–661. doi: 10.1109/COM-IT-CON54601.2022.9850820.

[73]. R. Khan, S. Hossain, A. Hossain, F. H. Siddiqui, and S. B. Noor, "Bangla Speech-Based Person Identification Using LSTM Networks," in Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST, 2023, pp. 358–370. doi: 10.1007/978-3-031-34619-4_29.

[74]. R. Naveen, C. Jeevan Reddy, R. Tanguturu, and M. Anand Kumar, "Speaker Identification and Verification using Deep Learning," in 2022 International Conference on Signal and Information Processing, IConSIP 2022, 2022. doi: 10.1109/ICoNSIP49665.2022.10007520.

[75]. M. Dua, C. Jain, and S. Kumar, "LSTM and CNN based ensemble approach for spoof detection task in automatic speaker verification systems," J Ambient Intell Humaniz Comput, vol. 13, no. 4, pp. 1985–2000, 2022, doi: 10.1007/s12652-021-02960-0.

[76]. Z. Ni et al., "Unusable spoken response detection with BLSTM neural networks," in 2018 11th International Symposium on Chinese Spoken Language Processing, ISCSLP 2018 - Proceedings, 2018, pp. 255–259. doi: 10.1109/ISCSLP.2018.8706635.

[77]. J. V Campos De Negreiros, C. Veiga Muniz, D. L. Dos Santos, F. R. R. Santos, M. G. Fernandes Costa, and C. F. F. Costa Filho, "Identification of Individuals Using Multimodal Data and LSTM Neural Networks," in International Conference on Electrical, Computer, Communications and Mechatronics Engineering, ICECCME 2023, 2023. doi: 10.1109/ICECCME57830.2023.10253325.

[78]. K. K. Katrak, K. Singh, A. Shah, R. Menon, and V. R. Badri Prasad, "Transformers for Speaker Recognition," in Smart Innovation, Systems and Technologies, 2022, pp. 49–62. doi: 10.1007/978-981-16-7996-4_5.

[79]. K. Khoria, A. T. Patil, and H. A. Patil, "On significance of constant-Q transform for pop noise detection," Comput Speech Lang, vol. 77, 2023, doi: 10.1016/j.csl.2022.101421.

[80]. M. V Evsyukov, M. M. Putyato, and A. S. Makaryan, "Antispoofing Countermeasures in Modern Voice Authentication Systems," in CEUR Workshop Proceedings. Yalta, Crimea, 2021, pp. 197–202.

[81]. M. R. Kamble and H. A. Patil, "Effectiveness of Mel scale-based ESA-IFCC features for classification of natural vs. spoofed speech," in Pattern Recognition and Machine Intelligence: 7th International Conference, PReMI 2017, Kolkata, India, December 5-8, 2017, Proceedings 7, Springer, 2017, pp. 308–316.

[82]. A. Babu, E. Raoul, G. Kassahun, I. Dufour, D. Mandal, and D. Thuau, "Programmable Polymeric-Interface for Voiceprint Biometrics," Adv Mater Technol, vol. 9, no. 4, p. 2301551, 2024.

[83]. T. Parlar, "Data privacy and security in the metaverse," in Metaverse: Technologies, Opportunities and Threats, Springer, 2023, pp. 123–133.

[84]. M. V Evsyukov, M. M. Putyato, and A. S. Makaryan, "Antispoofing Countermeasures in Modern Voice Authentication Systems," in CEUR Workshop Proceedings, 2021, pp. 197–202. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85122794382&partnerID=40&md5=6e038815dd07d23968182b493ff3a593

[85]. T. Parlar, "Data Privacy and Security in the Metaverse," in Studies in Big Data, vol. 133, 2023, pp. 123–133. doi: 10.1007/978-981-99-4641-9_8.

[86]. Y. Yanagi, R. Orihara, Y. Tahara, Y. Sei, T. Alumäe, and A. Ohsuga, "The Proposal of Countermeasures for DeepFake Voices on Social Media Considering Waveform and Text Embedding," Annals of Emerging Technologies in Computing, vol. 8, no. 2, pp. 15–31, 2024, doi: 10.33166/AETiC.2024.02.002.

AUTHOR CONTRIBUTIONS

Every author has contributed equally in this paper

CONFLICTS OF INTEREST

There is no conflict of interest

LIST OF AUTHORS

1. Pragati Jain Orcid Id:
0009-0000-8666-7129
2. Poorvi Ujjainia Orcid Id:
0009-0000-1170-4589
3. Anshika Srivastava Orcid Id:
0000-0003-0132-1060
4. Kajal Shrivastav Orcid Id:
0000-0002-2819-421X
5. Ishu Rani Orcid Id:
0009-0001-7889-5401
6. Akshat Vashisht Orcid Id:
0009-0005-2823-7952
7. Rudranarayan Behera Orcid Id:
0009-0004-7040-733X
8. Bhavika Moza Orcid Id:
0009-0001-9019-6267
9. Debhjit Mukherjee Orcid Id:
0009-0009-5673-2554