Grenze International Journal of Engineering and Technology, June Issue

# A Review on Nonvocal Password Recognition using Lip analysis Strategy

M. Senthil Kumar[1], A. Aafrin Nisha[2], R. Blesslinjaffy[3], A. Dhakshayani[4] and B. Chidambara Rajan[5]
[1-4]SRM Valliammai Engineering College/Department of Cyber Security, Chengalpattu, India
Email: msen1982@gmail.com, {afraanmiza, blesslinjaffy, ayamini2003}@gmail.com
[5]SRM Valliammai Engineering College/Department of Electronics and Communication Engineering, Chengalpattu, India
Email: profbcr@yahoo.com

***Abstract*— In the today's stage of increasing digital interactions and the need for secure authentication methods, traditional password-based systems face numerous challenges, such as susceptibility to theft, hacking, and the inconvenience of remembering complex passwords. In retaliation to these challenges, the paper presents a novel approach to authentication called the Silent Password, which proposed lip synchronization analysis and Convolution Neural Networks (CNNs) for user verification. The Silent Password system harnesses the unique characteristics of an individual's lip movements during speech, capitalizing on the fact that these movements are both distinct and difficult for impostors to replicate. Furthermore, the CNN model extracts spatial and temporal features from the lip movement data, which are then compared with the user's enrollment data. If the lip synchronization patterns match within an acceptable threshold, access is granted. This study demonstrates the feasibility of silent password as an innovative, secure and user-friendly authentication method, paving the way for future advancements in biometric-based user verification systems.**

***Index Terms*— Silent Password, Lip movement, Threshold, Pattern matching, Biometrics.**

## I. Introduction

In the area of information security, authentication holds a central and indispensable role. With the continuous evolution of digital landscapes, there is a growing demand for security but user-friendly methods of confirming identities. Traditional authentication methods, primarily dependent on alphanumeric passwords and Personal Identification Numbers (PINs), are loaded with vulnerabilities, spanning from human errors in password guidance to susceptibility to malicious attacks like brute force and phishing attempts. Faced with these challenges, biometric authentication methods have emerged as a budding solution for strengthen the security of digital systems while enhancing user ease. Among the diverse array of biometric authentication techniques, the evolving sector of soundless lip reading offers an intriguing and ingenious solution [15]. Soundless lip reading capitalizes on the unique patterns of lip movements and perceptible indication correlated with expressing language to discern the identities of individuals seeking access to secure systems or privileged data [16]. Unlike traditional static biometrics like fingerprints or facial features, this dynamic biometric technique necessitates the analysis of a user's lip movements as they silently articulate a predefined passphrase, providing the potential for robust yet modest verification [17]. This survey paper delves into a comprehensive exploration of the Silent Lip-Reading Password Analysis Project, an evolving application that investigates the use of silent lip reading as a biometric authentication

*Grenze ID: 01.GIJET.10.2.193*

method. Through the examination of users' lip movements as they silently pronounce a proposed passphrase, the project aims to develop a sophisticated soundless lip-reading password system that enhances security while maintaining user comfort [19]. The objective of this survey is to furnish a thorough and educative outline of the project's goals, methodologies, findings, and recommendations, illuminate on the ingenious progress being achieved in the domain of biometric authentication.

## II. LITERATURE SURVEY

The related works explores the design, implementation, and evaluation of the Silent Password system, highlighting its robustness against spoofing attempts and its potential for real-world applications in secure digital environments, such as investigating research on audio-visual speech recognition systems that combine lip movements and audio data for speech understanding. These systems often employ deep learning techniques, including CNNs and RNNs [3]. Then perform research on multi-modal biometric systems that combine different biometric modalities, such as face recognition, voice recognition, and lip movement analysis. These approaches can provide insights into the integration of lip synchronization with other authentication methods [5]. In addition, it provides techniques and technologies for detecting spoofing attempts in biometric systems [7]. Consider how the Silent Lip Password system can be designed to mitigate such attacks. Considerably exploring existing studies and systems that use lip movements for biometric authentication, it provides insights into the feasibility and challenges of using lip synchronization for password-based security.

[1] X. Ai and B. Fang have suggested that the incorporation of lip motion information into the language modeling process significantly improves the performance of automatic speech recognition and natural language understanding systems. The synchronized analysis of visual lip curves, acoustic features, and textual context provides a multi-modal framework that outperforms traditional unimodal approaches, particularly in noisy and challenging environments.

[2] G. Li et al has proposed that the integration of audio and visual signal from multiple channels enhances the quality of separated speech signals, effectively mitigates reverberation, and leads to superior speech recognition performance. The synergy between audio and visual data not only aids in source separation but also provides valuable contextual information that can disambiguate overlapping speech, making it a valuable resource for a wide range of applications.

[6] H. Wang et al have demonstrated that the 3D-CVT model outperforms traditional 2D-based methods, particularly in scenarios with varying lighting conditions, facial orientations, and occlusions. By capturing spatiotemporal features from lip movements in a three-dimensional context, the 3D-CVT model achieves superior accuracy in transcribing spoken words and phrases.

[7] Z. Pan et al have proposed that Selective Listening, in combination with lip synchronization, has the potential to significantly boost speech recognition accuracy, particularly when distinguishing and isolating individual speakers. By enabling the system to selectively attend to the speech and lip movements of a specific speaker, the paper shown promising results in mitigating background noise and enhancing the clarity of the recognized speech.

[8] D. Bigioi et al narrated implementation of the pipeline and shown that speech-driven facial landmark animation can produce highly realistic lip synchronization and facial expressions, enhancing the overall quality of automated dubbing. By accurately aligning facial landmarks with the spoken words and emotions of the dubbed content, the paper aims to potentially create engaging and lifelike animations for characters that speak different languages, capturing the nuances of their emotions and expressions more accurately.

[9] S. Bu et al explain the indications of T-F masks, when applied effectively, can effectively separate desired speech from background noise, leading to improved speech quality and intelligibility. By modeling the structure of the speech signal in the time-frequency domain, we have achieved a level of flexibility and adaptability that allows for effective noise reduction while preserving essential speech features. Additionally, contributes to the improvement of speech enhancement and recognition systems but also addresses critical challenges in real-world scenarios, such as noisy environments, multiple speakers, and non-stationary noise sources.

[10] E. Villalobos et al suggest that incorporating non-sensitive soft-biometric attributes can significantly reduce biases and disparities in face verification systems. By leveraging attributes such as age, gender, and ethnicity in a non-discriminatory manner, the paper promotes fairer and more equitable outcomes in face verification processes. Furthermore, the integration of fairness-aware machine learning techniques underscores the adaptability and potential of these models in ensuring fairness and equity in technology.

[11] L. Qin et al have narrated the subtle yet potent threat posed by low visual distortion and morphing attacks. These attacks, when executed with precision, can deceive facial recognition systems, leading to security breaches

and potential misuse of sensitive personal data. The ability to manipulate partial face images while maintaining a high degree of visual similarity poses a significant challenge for authentication and identity verification systems

[12] Q. Li and X. Lin provide the feasibility of implementing robust privacy-preserving mechanisms without compromising the accuracy and efficiency of speaker recognition. The successful development and integration of these techniques into 6G communication systems can empower individuals to engage with Cybertwins while maintaining their privacy and data security. The paper aims to set a crucial precedent for the development of privacy-respecting technologies that align with evolving ethical and regulatory considerations

[13] C. -Z. Yang et al have suggested that the significant challenge posed by Deep Fake attacks on speaker authentication systems. These attacks have the potential to undermine the integrity of voice-based security measures and pose serious risks to applications such as access control and financial transactions. The proposed dynamic lip analysis, as a countermeasure, has shown promise in identifying DeepFake attacks. By capturing the nuanced visual signal of genuine speech production, this approach serves as a valuable addition to speaker authentication systems. It introduces an additional layer of security, which can help protect against emerging threats in the evolving landscape of biometric security.

Native English-speaking colleague to proofread your paper. Do not add page numbers.

## III. Methodology

### A. Data Collection

In the silent lip password authentication system, the data collection involves the process of obtaining and recording video and audio information belonging to an individual's silent lip movements while they utter a specific set of password or passphrase [6]. This data is obtained with the purpose of developing a biometric authentication system that can authenticate a person's identity based upon the unique lip movement patterns correlated with silently pronounced a specific passphrase. Collect a dataset of individuals speaking a predefined passphrase or set of phrases. This dataset will serve as the reference data for the lip synchronization authentication system [1]. It should include a diverse group of users to account for individual variations. The collected data consists of video clippings of lip movements, audio recordings and other appropriate facial features, which serves as credential points for authentication. Then, this biometric data is examined and processed to build a reliable and unique method for user authentication based on silent lip recognition [3]. The data can be collected in following ways: It involves collecting video clippings of individuals uttering the password silently and record the corresponding audio, even though it is a silent lip analysis. This will serve as a reference point for syncing lip movement patterns with speech. The following data annotations is performed on each video recordings: Textual representation of the spoken passphrase, time stamps for specific sounds in the passphrase, analysis of lip movements [9] (key lip shapes, transition between shapes and lip trajectories) and some facial features such as facial landmarks, eye movements and some facial points. Lip regions from the video frames are extracted for preprocessing, normalizing these video frames and aligning those frames for compatible analysis and converting the audio frames to a specific format that will be used in synchronization with the video data [7]. The collected and processed data for password analysis is stored securely to prevent data breaches or any misuse.

### B. Video Capture

Capturing a video in image processing involves the acquisition of sequential frames from a video source, such as a camera device or a pre-recorded video file, and subsequently processing each frame individually for various tasks like object detection, tracking, segmentation, or any other form of analysis. The process typically consists of several steps: The first step is to initialize the video source [1]. This could be a camera device connected to the system or a pre-existing video file. Once the video source is initialized, frames are continuously acquired from the source. Each frame represents a single image at a specific instance in time. These frames are essentially arrays of pixel values representing the image. After acquiring each frame, various image processing techniques are applied to manipulate or extract information from the frame. This could involve a wide range of operations such as filtering, enhancement, edge detection, feature extraction, object detection, and recognition, among others. Optionally, the processed frames can be displayed in real-time to visualize the output of the image processing algorithms [7]. This step is particularly useful for debugging and monitoring the performance of the processing pipeline. Once all frames have been processed or when the user decides to stop the processing, the video source is released, and any allocated resources are freed. This ensures proper cleanup and prevents resource leaks. Silent lip password recognition involves using lip movements to authenticate users without the need for audible speech.

### C. Lip Feature Extraction

Analyzing lip features in image processing typically involves extracting relevant information from images or video frames that contain human lips. This process can serve various purposes such as lip reading, emotion recognition, speech recognition, or biometric identification. Here's a theoretical overview of how lip feature extraction is approached in image processing: Obtain images or video frames containing human faces, ensuring that the lips are clearly visible and well-defined. Use face detection algorithms to localize the face region within the image. This step helps in focusing the subsequent analysis specifically on the lip region [12]. Perform preprocessing techniques such as histogram equalization, noise reduction, or contrast enhancement to improve the quality and clarity of the lip region. Utilize techniques to accurately locate the mouth region within the detected face. This step is crucial for isolating the lips from other facial features. Employ segmentation algorithms to separate the lip region from the rest of the mouth area and surrounding facial features. This could involve color-based segmentation, edge detection, or model-based segmentation. Extract geometric features such as lip contour, lip width, lip height, and lip curvature. These features provide information about the overall shape and structure of the lips. Analyze textural features within the lip region, such as wrinkles, creases, or skin texture patterns. Texture analysis can provide additional discriminative information for lip characterization. Extract color-based features from the lip region, such as average lip color, lip hue, saturation, and brightness [15]. Color analysis can be particularly useful for tasks like emotion recognition or detecting lipstick colors. Track the movement of the lips across consecutive video frames to capture temporal dynamics. This helps in analyzing lip motion patterns, which are essential for tasks like lip reading or speech recognition. Extract dynamic features such as lip motion velocity, lip opening/closing speed, or lip shape changes over time. These features provide insights into speech articulation and phoneme production. Combine the extracted features into a feature vector representation for each lip instance or frame. This feature vector serves as input for subsequent analysis tasks such as classification, regression, or clustering. Optionally, apply dimensionality reduction techniques such as PCA (Principal Component Analysis) or LDA (Linear Discriminant Analysis) to reduce the dimensionality of the feature space while preserving discriminative information [20]. Utilize the extracted lip features for recognizing spoken words or phonemes by mapping them to corresponding linguistic units. Overall, the process of lip feature extraction in image processing involves a combination of preprocessing, localization, feature extraction, temporal analysis (for videos), and application-specific analysis tailored to the task or application at hand [17]. Advanced techniques from computer vision, image processing, and machine learning are often employed to achieve accurate and robust lip feature analysis.

*D. Feature Representation*

Convert the extracted lip features into a numerical representation that can be used for comparison and authentication. Common methods include image processing techniques and computer vision algorithms to create a feature vector [13]. In silent password recognition, feature representation plays a crucial role in transforming raw input data into a format that is suitable for machine learning algorithms to analyze and make decisions. Silent password recognition typically involves identifying users based on their unique behavioral or physiological characteristics, such as typing patterns, mouse movements, or, in this case, lip movements or gestures performed silently [3]. Users are required to perform a predefined set of silent lip movements or gestures. These movements can include opening and closing the lips in specific patterns, forming shapes with the lips, or other gestures that can be reliably captured by a camera. Raw data acquired from the lip movements may need preprocessing to enhance its quality and remove noise. This could involve techniques such as noise reduction, illumination normalization, or background subtraction to isolate the lip region. Feature extraction involves identifying and quantifying key characteristics of the lip movements that are discriminative and invariant across users. Features could include: Extract geometric properties of the lip movements, such as lip contour, lip area, lip width. Capture temporal dynamics of the lip movements, such as velocity, acceleration, or duration of specific gestures. Analyze textural patterns within the lip region, such as wrinkles, creases, or skin texture, using techniques like local binary patterns (LBP) or Gabor filters. Describe the spatial arrangement of key points on the lip contour using techniques like shape context descriptors. Calculate optical flow to capture the motion of lip pixels between consecutive frames, providing information about the direction and speed of lip movement [4]. Extract features from pre-trained deep learning models (e.g., CNNs) applied to lip images or sequences, leveraging learned representations. After extracting a set of relevant features from the lip movements, these features are represented in a structured format suitable for machine learning algorithms. Concatenating extracted features into a single feature vector for each lip movement instance. Scaling or normalizing feature values to ensure consistency and comparability across different features. Optionally, applying dimensionality reduction techniques such as PCA or autoencoders to reduce the dimensionality of the feature space while preserving essential information. For temporal data, representing sequences of feature vectors using methods like temporal pooling or recurrent neural networks (RNNs) to capture temporal dependencies. Finally, the feature representations are used as input to machine learning models or pattern recognition algorithms for training and recognition tasks [10]. These models can include classifiers (e.g., SVM, Random Forest), neural networks or more specialized models tailored to the specific characteristics of silent lip movements. In summary, feature representation in silent password recognition involves extracting relevant features from lip movements, transforming them into a structured representation, and using this representation for user identification or authentication purposes. Effective feature representation is crucial for capturing the unique characteristics of users' silent lip movements while minimizing intra-class variations and maximizing inter-class differences.

*E. Enrollment Phase*

During the enrollment phase, users record themselves speaking the passphrase multiple times [11]. The system extracts and stores the lip features and their corresponding feature vectors for each user. The enrollment phase in silent password recognition is a critical step where users' unique behavioral or physiological characteristics, in this case, their silent lip movements or gestures, are captured and stored in the system for subsequent authentication or identification purposes. The enrollment process involves several key steps: Users are registered into the system by providing necessary identification information (e.g., username, email) and possibly undergoing initial authentication steps (e.g., password, biometric verification) to ensure the legitimacy of the user. Users are instructed on the specific silent lip movements or gestures required for authentication. This may involve demonstrating the required movements using visual aids or multimedia instructions to ensure users understand the task. During the enrollment phase, users perform the predefined set of silent lip movements or gestures multiple times while being recorded by a camera or sensor. Each session captures a sequence of lip movements performed by the user. Raw data acquired from the lip movements may undergo preprocessing to enhance its quality and remove noise. This could include techniques such as noise reduction, illumination normalization, or background subtraction to isolate the lip region. Features representing the unique characteristics of users' silent lip movements are extracted from the preprocessed data [12]. These features could include geometric properties, temporal dynamics, texture patterns, or deep learning representations of the lip movements. The extracted features are structured and represented in a format suitable for storage and subsequent comparison during authentication. This representation may involve creating feature vectors or sequences of feature vectors that capture the essence of users' lip movements. Based on the extracted features, enrollment templates are created for each registered user. These templates contain information about the user's unique lip movement characteristics and serve as reference points during authentication. The enrollment templates, along with any necessary user metadata, are securely stored in the system's database or storage repository. It's essential to employ robust security measures to protect users' sensitive biometric information. Optionally, the quality of enrolled templates may be assessed to ensure they meet predefined criteria for reliability and consistency [2]. This could involve evaluating the variability of lip movements across enrollment sessions or comparing features against predefined thresholds. Users may receive feedback on the success of the enrollment process, confirming their successful registration into the system. This feedback could include visual indicators, messages, or alerts indicating successful completion. The enrollment phase establishes a baseline for users' silent lip movements within the recognition system, enabling subsequent authentication or identification based on comparisons between stored enrollment templates and real-time lip movements during authentication attempts. A robust and accurate enrollment process is crucial for ensuring reliable and secure silent password recognition.

*F. Authentication Phase*

When a user attempts to access a system or perform a transaction, they are prompted to speak the passphrase silently [12]. The system captures their lip movements and extracts the lip features. The authentication phase in silent password recognition is the process by which a user's identity is verified based on their previously enrolled silent lip movements or gestures. This phase involves comparing the lip movements performed by the user during authentication against the stored enrollment templates to determine whether the user's identity can be confirmed. Here's a theoretical explanation of the authentication phase: The authentication process begins when a user initiates an authentication attempt, typically by providing their username or other identifying information. During the authentication phase, the user performs the predefined set of silent lip movements or gestures while being recorded by a camera or sensor. The lip movements performed by the user are captured as a sequence of frames. Raw data acquired from the lip movements may undergo preprocessing to enhance its quality and remove noise. Similar preprocessing steps applied during the enrollment phase may be employed to ensure consistency. Features representing the user's lip movements are extracted from the preprocessed data. These features capture the unique characteristics of the user's lip movements and are typically derived using techniques such as geometric analysis, texture analysis, or deep learning-based representations. The extracted features are structured and represented in a format suitable for comparison against the stored enrollment templates. This representation may involve creating feature vectors or sequences of feature vectors that capture the essence of the user's lip movements during authentication [3]. The extracted features from the authentication attempt are compared against the stored enrollment templates for the corresponding user. Various similarity measures or classification algorithms may be employed to assess the similarity between the authentication features and the enrolled templates. Based on the comparison results, a decision is made regarding the user's identity. If the extracted features closely match the enrollment templates within an acceptable threshold, the user's identity is authenticated, and access is granted. Otherwise, the authentication attempt may be rejected.

The authentication phase verifies the user's identity based on their silent lip movements, enabling secure and convenient access to the system without the need for traditional passwords or physical tokens [16]. A reliable and accurate authentication process is essential for maintaining the security and usability of the silent password recognition system.

*G. Feature Comparison*

Feature comparison in silent password recognition involves the process of comparing the features extracted from a user's lip movements during authentication with the features stored in the enrollment templates to determine the similarity or dissimilarity between them. This comparison is crucial for verifying the identity of the user and granting access to the system. During the authentication phase, features representing the user's lip movements are extracted from the raw data captured during the authentication attempt. These features capture various characteristics of the lip movements, such as geometric properties, temporal dynamics, or texture patterns [4]. The extracted features are structured and represented in a format suitable for comparison. This representation may involve creating feature vectors or sequences of feature vectors that encapsulate the relevant information extracted from the lip movements. The enrollment templates corresponding to the user attempting authentication are retrieved from the system's database or storage repository. These templates contain the features extracted during the enrollment phase and serve as reference points for comparison. Once the enrollment templates and authentication features are obtained, a comparison is performed to assess their similarity or dissimilarity. Analyzing the correlation coefficients between corresponding features in the enrollment templates and authentication features. A decision threshold or acceptance criterion is defined to determine whether the features extracted during authentication sufficiently match the enrollment templates to authenticate the user [12]. This threshold may be based on empirical observations, statistical analysis, or domain-specific requirements. Based on the comparison results and the defined threshold, a decision is made regarding the user's identity. If the similarity between the authentication features and enrollment templates exceeds the threshold, the user's identity is authenticated, and access is granted. Otherwise, the authentication attempt may be rejected. Users may receive feedback on the outcome of the authentication attempt, confirming whether their identity has been successfully authenticated or indicating any authentication failures. This feedback could include visual indicators, messages, or alerts. Robust security measures are employed to protect the feature comparison process and ensure the integrity of users' biometric information [16]. This may include encryption of data transmission, secure storage of enrollment templates, and mechanisms to prevent spoofing attacks. Feature comparison forms the core of the authentication process in silent password recognition, enabling reliable and secure verification of users' identities based on their unique lip movement characteristics. A robust comparison mechanism is essential for maintaining the security and accuracy of the authentication system.

*H. User Authentication*

User authentication in silent password recognition is the process of verifying the identity of a user based on their previously enrolled silent lip movements or gestures. This process involves several steps to ensure the security and accuracy of the authentication [6]. Here's a theoretical explanation of user authentication in silent password recognition. The authentication process begins when a user initiates an authentication attempt, typically by providing their username or other identifying information. During the authentication attempt, the user performs the predefined set of silent lip movements or gestures while being recorded by a camera or sensor. The lip movements performed by the user during the authentication attempt are captured as a sequence of frames or data points. Raw data acquired from the lip movements may undergo preprocessing to enhance its quality and remove noise. This could include techniques such as noise reduction, illumination normalization, or background subtraction [14]. Features representing the user's lip movements are extracted from the preprocessed data. These features capture the unique characteristics of the user's lip movements and are typically derived using techniques such as geometric analysis, texture analysis, or deep learning-based representations. The extracted features from the authentication attempt are compared against the stored enrollment templates for the corresponding user.

Various similarity measures or classification algorithms may be employed to assess the similarity between the authentication features and the enrolled templates. A decision threshold or acceptance criterion is defined to determine whether the features extracted during authentication sufficiently match the enrollment templates to authenticate the user. This threshold may be based on empirical observations, statistical analysis, or domain-specific requirements. Based on the comparison results and the defined threshold, a decision is made regarding the user's identity [17]. If the similarity between the authentication features and enrollment templates exceeds the threshold, the user's identity is authenticated, and access is granted. Otherwise, the authentication attempt may be rejected. Robust security measures are employed to protect the authentication process and ensure the integrity of users' biometric information. This may include encryption of data transmission, secure storage of enrollment templates, and mechanisms to prevent spoofing attacks. User authentication in silent password recognition aims to provide a secure and convenient method for verifying users' identities based on their unique lip movement characteristics. A reliable and accurate authentication process is essential for maintaining the security and usability of the silent password recognition system.

## IV. Discussions

This paper offers a substantial survey of lip biometrics research, particularly focusing on lip integration for authentication applications. It explores a range of techniques and assesses their performance. The foundation for processing visual data and capturing temporal information from lip movements is built upon Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN). Firstly, or video clippings lip figures undergo preprocessing to strengthen the input data quality. This can encompass the functions like facial feature perception and arrangement to detain the visible patterns, lip contours, and surface. This technique is a prevalent approach to construct hybrid planning that merge CNNs and RNNs [8]. CNNs draw out the spatial characteristics from an individual frames, while RNNs sequentially process these features to represent temporal dynamics. Consequently, 3D CNNs can be deployed to directly handle spatiotemporal data from video, where the third dimension signifies time. This facilitates the incorporation of spatial and temporal information. Though, this method launches data enhancement techniques, including random flipping, jittering, and cropping, to enhance the model's resilience to variations in the appearance of lip movement. The approach mainly discusses the capturing and analyzing the patterns to create a unique identifier for authentication purposes. Unlike some biometric methods, lip synchronization does not require physical contact. It can be captured through video or imaging, making it a non-intrusive form of authentication. One of the primary challenges is achieving a high level of accuracy [10]. Lip movements can be affected by various factors such as lighting conditions, facial hair, and changes in facial expressions, which can impact the reliability of the authentication system. As with any biometric system, there is a concern about spoofing or impersonation.  Users may have concerns about the privacy implications of capturing and storing biometric data, even if it's based on lip movements. To overcome these problems, the proposed methodology, implement robust security measures to protect this sensitive information. Lip movements can vary across cultures and individuals. Ensuring that the system is inclusive and works well for a diverse user base is an important consideration and provides seamless integration of lip synchronization into the user experience. Finally, data protection and privacy laws has been taken into consideration to ensure that the collection, storage, and processing of biometric data comply with regulations in different regions. The design of Convolutional Neural Network (CNN) architecture for a soundless password system based on lip synchronization involves creating a model that can effectively extract features from lip movements and perform classification or authentication tasks that Crop and resize video frames, normalize pixel values, and pre-process the data to enhance the model's ability to learn and Depending on the task, the output layer for binary or multi-class classification is designed [5]. Additionally, apply batch normalization to stabilize and accelerate training to introduce dropout layers to prevent overfitting. It is discussed that Training model works by splitting the dataset into training and testing sets then apply data augmentation techniques to increase the diversity of the training set and Train the model on the training set using the compiled model and monitor performance on the validation set. Furthermore, the approach evaluates the model on the testing set to assess its generalization performance and Use chosen evaluation metrics to measure performance.

## V. Conclusions

This study has developed a contemporary and promising strategy for biometric authentication, which involves a soundless password system based on lip synchronization. These systems capitalize on the unique lip movement patterns of individuals to verify their identity, providing a user-oriented and probably highly assured replacement to conventional password-based techniques. Therefore, individuals exhibit distinctive lip motion patterns even when silently articulating passphrases or phrase. This uniqueness serves as the principle for employing lip integration as a biometric validation characteristic, which presents a user-oriented lip integration-based validation system. This system eliminates the need for users to remember complex passwords or PINs. Integrating the silent lip integration system with other biometric components can far improvise the security level in the ever-evolving information system. The paper conclude that lip analysis offers a compelling alternative to traditional authentication methods, capitalizing on the uniqueness of everyone's lip movements. This innovation combines security and user convenience in a manner that is poised to redefine the future of digital access. Furthermore, it holds great potential in security while eliminating the inconvenience of traditional passwords such as PINs, passwords etc. This promising technology offers a glimpse into the future of secure and seamless user authentication. The integration of lip analysis into authentication systems represents a ground-breaking approach that not only enhances security but also streamlines the user experience. As technology continues to evolve, silent passwords based on lip analysis are well-positioned to play a central role in the field of digital security. Finally embracing this technology means not just securing our systems but also simplifying the user experience in our increasingly digital world. The accuracy enhancement through research and development efforts should focus on improving the accuracy of lip analysis for authentication. This includes refining algorithms, increasing the specificity of lip movement detection, and reducing false acceptance and rejection rates. This can help throughout spoofing attempts [18]. Also, explore the integration of lip analysis with other biometric or authentication methods, such as facial recognition, fingerprint scanning, or behavioural biometrics, to create more robust multi-factor authentication systems.in future this method conducts extensive real-world testing of lip analysis systems in various environments and conditions to assess their robustness and reliability and address privacy and ethical concerns associated with the collection and storage of biometric data, including compliance with data protection regulation.

## REFERENCES


[1] X. Ai and B. Fang, "Cross-Modal Language Modeling in Multi-Motion-Informed Context for Lip Reading," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 31, pp. 2220-2232,2023,doi:10.1109/TASLP.2023.3282109.

[2] G. Li et al., "Audio-Visual End-to-End Multi-Channel Speech Separation, Dereverberation and Recognition," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 31, pp. 2707-2723, 2023, doi: 10.1109/TASLP.2023.3294705.

[3] S. S. Mahdi et al., "Matching 3D Facial Shape to Demographic Properties by Geometric Metric Learning: A Part-Based Approach," in IEEE Transactions on Biometrics, Behavior, and Identity Science, vol. 4, no. 2, pp. 163-172, April 2022, doi: 10.1109/TBIOM.2021.3092564.

[4] H. X. Pham, Y. Wang and V. Pavlovic, "Learning Continuous Facial Actions From Speech for Real-Time Animation," in IEEE Transactions on Affective Computing, vol. 13, no. 3, pp. 1567-1580, 1 July-Sept. 2022, doi: 10.1109/TAFFC.2020.3022017.

[5] H. Liu, H. Cai, Q. Lin, X. Li and H. Xiao, "Adaptive Multilayer Perceptual Attention Network for Facial Expression Recognition," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 32, no. 9, pp. 6253-6266, Sept. 2022, doi: 10.1109/TCSVT.2022.3165321.

[6] H. Wang, G. Pu and T. Chen, "A Lip Reading Method Based on 3D Convolutional Vision Transformer," in IEEE Access, vol. 10, pp. 77205-77212, 2022, doi: 10.1109/ACCESS.2022.3193231.

[7] Z. Pan, R. Tao, C. Xu and H. Li, "Selective Listening by Synchronizing Speech with Lips," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 30, pp. 1650-1664, 2022, doi: 10.1109/TASLP.2022.3153258.

[8] D. Bigioi, H. Jordan, R. Jain, R. McDonnell and P. Corcoran, "Pose-Aware Speech Driven Facial Landmark Animation Pipeline for Automated Dubbing," in IEEE Access, vol. 10, pp. 133357-133369, 2022, doi: 10.1109/ACCESS.2022.3231137.

[9] S. Bu, Y. Zhao, T. Zhao, S. Wang and M. Han, "Modeling Speech Structure to Improve T-F Masks for Speech Enhancement and Recognition," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 30, pp. 2705-2715, 2022, doi: 10.1109/TASLP.2022.3196168.

[10] E. Villalobos, D. Mery and K. Bowyer, "Fair Face Verification by Using Non-Sensitive Soft-Biometric Attributes," in IEEE Access, vol. 10, pp. 30168-30179, 2022, doi: 10.1109/ACCESS.2022.3158967.

[11] L. Qin, F. Peng, S. Venkatesh, R. Ramachandra, M. Long and C. Busch, "Low Visual Distortion and Robust Morphing Attacks Based on Partial Face Image Manipulation," in IEEE Transactions on Biometrics, Behavior, and Identity Science, vol. 3, no. 1, pp. 72-88, Jan. 2021, doi: 10.1109/TBIOM.2020.3022007.

[12] Q. Li and X. Lin, "Efficient and Privacy-Preserving Speaker Recognition for Cybertwin-Driven 6G," in IEEE Internet of Things Journal, vol. 8, no. 22, pp. 16195-16206, 15 Nov.15, 2021, doi: 10.1109/JIOT.2021.3097266.

[13] C. -Z. Yang, J. Ma, S. Wang and A. W. -C. Liew, "Preventing DeepFake Attacks on Speaker Authentication by Dynamic Lip Movement Analysis," in IEEE Transactions on Information Forensics and Security, vol. 16, pp. 1841-1854, 2021, doi: 10.1109/TIFS.2020.3045937.

[14] B. -F. Wu, B. -R. Chen and C. -F. Hsu, "Design of a Facial Landmark Detection System Using a Dynamic Optical Flow Approach," in IEEE Access, vol. 9, pp. 68737-68745, 2021, doi: 10.1109/ACCESS.2021.3077479.
[15] M. Ezz, A. M. Mostafa and A. A. Nasr, "A Silent Password Recognition Framework Based on Lip Analysis," in IEEE Access, vol. 8, pp. 55354-55371, 2020, doi: 10.1109/ACCESS.2020.2982359.
[16] R. Tolosana, R. Vera-Rodriguez, J. Fierrez and J. Ortega-Garcia, "BioTouchPass2: Touchscreen Password Biometrics Using Time-Aligned Recurrent Neural Networks," in IEEE Transactions on Information Forensics and Security, vol. 15, pp. 2616-2628, 2020, doi: 10.1109/TIFS.2020.2973832.
[17] J. -W. Kim, H. Yoon and H. -Y. Jung, "Linguistic-Coupled Age-to-Age Voice Translation to Improve Speech Recognition Performance in Real Environments," in IEEE Access, vol. 9, pp. 136476-136486, 2021, doi: 10.1109/ACCESS.2021.3115608.
[18] W. Mu and B. Liu, "Voice Activity Detection Optimized by Adaptive Attention Span Transformer," in IEEE Access, vol. 11, pp. 31238-31243, 2023, doi: 10.1109/ACCESS.2023.3262518.
[19] W. Gao et al., "Digital Retina: A Way to Make the City Brain More Efficient by Visual Coding," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 31, no. 11, pp. 4147-4161, Nov. 2021, doi: 10.1109/TCSVT.2021.3104305.
[20] M. Szymkowski, E. Saeed, M. Omieljanowicz, A. Omieljanowicz, K. Saeed and Z. Mariak, "A Novelty Approach to Retina Diagnosing Using Biometric Techniques with SVM and Clustering Algorithms," in IEEE Access, vol. 8, pp. 125849-125862, 2020, doi: 10.1109/ACCESS.2020.3007656.