International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

Survey Analysis on Secured user Authentication through Biometric Recognition

Preethi S[1], Joseph Charles P[2]
[1]DepartmentofComputerScience,St. Joseph's College (A), Bharathidasan University, Trichy, India

Email:preethidiwakarmphil@gmail.com

[2]DepartmentofInformationTechnology, St. Joseph's College (A), Bharathidasan University, Trichy, India

Email: hedrpjcharles@gmail.com

Abstract

Secured user authentication is the process of verifying the user authenticity. Biometric authentication is the human identification system employed to match the biometric characteristics of user for verifying the authenticity. Biometric identifiers are exclusive, making it harder to hack accounts using them. Common types of biometrics comprise the fingerprint scanning verifies authentication based on a user's fingerprints Face recognition and voice recognition are employed in real-time application for improving the security level in different application scenarios. Face recognition is a method of identifying or verifying the individual identity using their face expression. Voice recognition is the ability of machine to receive and interpret the dictation to understand. Many researchers carried out their research on different face and voice recognition methods. But, recognition accuracy was not improved with minimum time consumption by existing biometric recognition method. In this research, different recognition methods are reviewed using biometric recognition method for user authentication. The recognition methods are efficiently on human faces dataset with respect to performance metrics like recognition accuracy, error rate, and recognition time.

Keywords: Secured User Authentication, Biometric Authentication, Face Recognition, Voice Recognition

1. Introduction

Biometric is used to recognize the identity of people when compared to template based on their particular characteristics. Biometric is a technological and scientific authentication based on biology and information assurance (IA). Biometrics refers to the human identification by their characteristics or traits. Biometric systems present the solution to guarantee that rendered services are accessed only by legitimate user. Biometric recognition refers the distinctive physiological features like voice, fingerprints, face, retina, iris and behavioral features like gait, signature characteristics called biometric identifiers for identifying and verifying the human identity. Biometric Authentication is the process of establishing or confirming the user as an authentic one. Biometric authentication has grown in popularity to provide personal identification. Fingerprint recognition is the process of ascertaining the uniqueness of single-centered on impressions. Face is employed as a biometric recognition because face traits differ from person to person.

This paper is arranged as: Section 2 reviews the drawbacks on existing face and voice recognition techniques. Section 3 shows the study and analysis of existing face and voice recognition methods. Section 4 describes the possible comparison between them. Section 5 explains the limitations of existing face and voice recognition techniques. The paper conclusion is provided in the section 6.

2. Literature Review

A modified technology acceptance model (TAM) was developed in Wang et al. [1] by including two new factors such as perceived privacy and perceived trust to the verify user recognition of biometric identification. But the designed model did not categorize the user through demographic segmentation with typical structural equation modeling study, thus possibly affecting the generalizability of the results. A spiking neural network (SNN) approach was introduced in Mansouri-Benssassi and Ye [2] for predicting emotional states based on

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

the facial expression and speech data. But, the designed approach was not used for continuous multi-modal emotion recognition to multi-sensory integration. The emotion recognition accuracy was not at required level. A new multimodal emotion recognition approach was introduced in Lee et al. [3] to increase BERT model performance for emotion recognition. The designed approach joined the fine-grained representation of audio and visual features with fine-tuning modalities. But, the designed approach increased the computation cost because of trainable weight and hyper parameters. The dynamic information was employed in Lavan et al. [4] for informing matching decisions. The dynamic stimuli were employed with shared information across modalities in articulatory mouth movements. However, the face-voice matching was not possible under circumstances with chance performance for dynamic face-voice matching. An end-to-end speech driven facial animation algorithm was introduced in Chai et al. [5] to create the lip-synchronized facial animation from vocal audio clip. A light- weight speech encoder was employed with robust vocal features from input audio. But, designed mechanism did not have ability to control the emotion or speaking style for enhancing expressiveness of facial animation. The second dysarthric-specific automatic speech recognition (ASR) system called Speech Vision (SV) was employed in Shahamiri [6] with dysarthric ASR challenges. SV employed visual data augmentation and generated the synthetic dysarthric acoustic visuals. But, synthetic data generation attained same output for same prompt that generate one additional sample per word for each speaker. A bimodal fusion algorithm was introduced in Wang et al. [7] to realize the speech emotion recognition. MFCC were employed to switch the speech signal into images. The weighted decision fusion method was used for combining the facial expression and speech signal to attain the speech emotion recognition. But, the recognition time was not reduced by bimodal fusion algorithm. A holistic solution was attained in Saxena and Varshney [8] for Smart Home Security for improving privacy and security using facial authentication and speech recognition. The designed method performed facial recognition through real-time feed of person at door. However, the masked users were not recognized solely on the basis of their eye region. A cancelable biometric template protection scheme was performed in Bansal and Garg [9] to base on format-preserving encryption and Bloom filters. The bloom filters aids to improve the security of input biometric template and recognition with better recognition performance. In Tse and Hung [10], user behavioral biometric identification using a multimodal scheme was introduced into utilizes the keystroke biometrics for protected and dependable mobile device identification. However, the biometric identification mechanisms have limited accuracy and efficacy. A biometric-enabled and Hyperledger fabric-based architectural framework was proposed in Faruk et al. [11] to e-Voting applications to automate identity verification. The security level was improved. In Pereira et al. [12], an electrocardiogram data acquisition method was employed into data acquisition protocol of an ECG signal in the biometric identification process. Biometric Security System was performed in Almomani et al. [13] to hybridizing auto-encoder (AE) network and a chaos-based ciphering algorithm to cipher the details of the stored biometric patterns and ensures their secrecy. A Secure Cancellable Biometric Cryptosystem was applied in Elazm et al. [14] to 3D chaotic cryptosystem. The rationale behind the utilization of the 3D chaotic cryptosystem is to guarantee strong encryption of biometric templates, and improve the security and privacy of users. A novel encoding technique was utilized in Harikrishnan et al. [15] to an innovative model which generates a One Time Iris Code (OTIC) for every user authentication. However, minimize the possibility of compromising the original Iris Code.

The contribution of this paper follows as,

- A novel multimodal emotion recognition approach was introduced to improve the BERT model performance for emotion recognition.
- An automated integrated speech signal and facial image analysis system was employed for identification of human emotions with better classification accuracy.
- An end-to-end speech driven facial animation algorithm was performed into generate lip-synchronized facial animation from vocal audio clip.
- A fully differentiable neural network model termed Taris was proposed into decoding the audio-only and audio-visual speech in real time applications.
- A novel multimodal fusion attention network was introduced for audio-visual emotion recognition depending on adaptive and multi-level factorized bilinear pooling (FBP).

A bimodal fusion algorithm was introduced to recognize the speech emotion recognition for facial expression and speech information.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: 25 July 2023 Revised: 12 September 2023 Accepted: 30 September 2023

3. Secured User Authentication Through Biometric Recognition

Biometric recognition is an automated information system for person or group of people identification based on physiological and behavioral characteristics. Biometric recognition provides the access control to the physical facilities and financial accounts for increasing the access efficiency to services and their utilization. Biometric face recognition is the ability of a biometric machine to identify and recognize the individual face for granting the access to secured system. The face recognition finds the person details through face matching with data in machine system. Fingerprint recognition is an automated method for identifying the individual identity depending on fingerprint comparison.

3.1 Multimodal Emotion Recognition Fusion Analysis Adapting BERT with Heterogeneous Feature Unification

A new multimodal emotion recognition approach was introduced to enhance the BERT model performance for emotion recognition. The designed approach used heterogeneous features depending on the language, audio and visual modalities. Self-Multi-Attention Fusion module, multi-Attention fusion module and Video Fusion module were joined with multimodal fusion mechanism to construct the transformer architecture. The fine-grained representation of audio and visual features was joined into common embedding for fine-tuning. Heterogeneous Features Unification (HFU-BERT) with BERT combined the heterogeneous features extracted from handcrafted and deep learning techniques for accurately predicting emotions. BERT extracted textual features with other modality-based features through structuring network with suitable emphasis positioned on every feature modality. The information was gathered from different unimodal features with relevant saliency and joined with suitable placement of relative weights among modalities for precisely recognizing the emotions.

3.2 An automated integrated speech and face image analysis system for the identification of human emotions

An automated integrated speech signal and facial image analysis system was designed for identification of human emotions like Normal (N), Happy (H), Sad (S), Disgust (D), Fear (F), Anger (A) and Surprise (Su). The multi-classification analysis was carried out to choose the relevant features for identifying different human emotions. The selected features or feature combination with age and gender were used to construct the learning-based classifiers with higher classification accuracy. The statistical speech and face image features were employed to differentiate between groups like N vs H, N vs S, N vs D, N vs F, N vs A, N vs Su.

The speech and face image features were obtained from the continuous speech and face images. Features were used to find different emotional human states. An integrated system was used for emotional state identification from automatic analysis of free speech and image face analysis.

3.3 Speech-Driven Facial Animation with Spectral Gathering and Temporal Attention An end-to-end speech driven facial

Animation algorithm was introduced to generate lip-synchronized facial animation from vocal audio clip. A light-weight speech encoder was employed to study the useful and robust vocal features from input audio without resorting to pre-trained speech recognition modules or large training data. The designed encoder was combined the spectral-dimensional bidirectional long short-term memory and temporal attention mechanism. The deformation gradients were employed as internal representation to perform nuanced local motions synthesized through vertex offsets. A lightweight, robust speech encoder was employed for animating 3D face avatars from input vocal audio to learn the subject-independent facial motion. The deformation gradients were employed as motion representation for non-rigidity handling and generalization to different faces. A spectral-dimensional bidirectional long short-term memory was used to develop the formant features spanning the wide spectrogram. Frame- wise attention mechanism was used to synchronize the lip motions with vocal signals in temporal dimension.

3.4 Taris: An online speech recognition framework with sequence-to-sequence neural networks for both audio-only and audio-visual speech

A fully differentiable neural network model termed Taris was introduced for decoding the audio-only and audio-visual speech in real time applications. AV Align and Taris were introduced for audio-visual speech integration and online speech recognition correspondingly. AV Taris was superior to audio-only variant of Taris for describing the visual modality utility to speech recognition within real time decoding

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

framework. Taris provided fully differentiable training pipeline. AV Taris has the potential to popularize the Audio-Visual Speech Recognition (AVSR) technology and addressed the inherent limitations of audio modality with less optimal listening conditions. A multimodal extension of Taris system attained fully differentiable solution to online audio-visual speech recognition. The cross-modal attention mechanism was employed in AV Align through limiting the attention span to a fixed window of video representations on every audio frame. Taris was used for sequence-to-sequence model over the consecutive audio windows. An additional end-of-block token in output domain was emitted once per every audio window. Taris eliminated the issues without using end-of-block tokens. Taris examined the dynamic windows of speech on word of interest. But the Taris was not assured the reliable segmentation of the spoken utterance into words.

3.5 Information Fusion in Attention Networks Using Adaptive and Multi-level Factorized Bilinear Pooling for Audio-visual Emotion Recognition

A novel multimodal fusion attention network was introduced for audio-visual emotion recognition depending on adaptive and multi-level factorized bilinear pooling (FBP). The designed framework linked temporal feature sequence to single label. A fully convolutional network (FCN) was used with 1-D attention mechanism and local response for speech emotion recognition. A global FBP (G-FBP) approach was designed to perform audio-visual information fusion through combining self-attention based video stream with audio stream. An adaptive global strategy termed AG-FBP was used to determine the fusion weight of modalities based on the emotion-related representation vectors from attention mechanism of respective modalities. An adaptive and multi- level FBP (AM-FBP) was introduced with global-trunk and intra-trunk data in one recording on AG-FBP to use the local emotion information. 1-D attention-based decoder was introduced to attain information related to emotion after audio encoder and video encoder respectively. An adaptive and multi-level FBP approach was introduced for audio-visual fusion. Depending on the fusion vector, output posterior probabilities of emotion classes was generated through fully-connected (FC) layer and softmax layer.

3.6 Human emotion recognition by optimally fusing facial expression and speech feature

A bimodal fusion algorithm was introduced to recognize the speech emotion recognition for facial expression and speech information. Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) were obtained to attain the facial emotion recognition. Mel Frequency Cepstrum Coefficient was used to convert speech signal into images. Long short-term memory (LSTM) and CNN was used to identify the speech emotion. The weighted decision fusion method was used for integrating the facial expression and speech signal to attain the speech emotion recognition. CNN architecture was used to extract the deep feature of an image sequence.The features output from the fully connection layer as the deep representation. RNN architecture learned the relationship among input deep representation at learning time related sequences. RNN architecture comprised the hidden layer and Dropout layer was inserted into hidden layer to eliminate overfitting issues.

4. Performance Analysis of Secured User Authentication through Biometric Recognition

Experimental evaluation of existing secured user authentication techniques is implemented. During the result analysis, the number of facial images is considered as input from Human Faces. The URL of the dataset is given as https://www.kaggle.com/datasets/ashwingupta3012/human-faces. The main aim of the dataset is a web scraped dataset of human faces suggested for image processing models.

5. Result and Discussion

The Result analysis of proposed method using different existing methods namely Automated integrated speech signal and facial image analysis system, End-to-end speech driven facial animation algorithm, Taris Model, Novel multimodal fusion attention network and Bimodal fusion algorithm.

Result analysis are carried out with existing methods with parameters are,

- Recognition Accuracy,
- Error Rate and
- Recognition Time

5.1 Analysis on Recognition Accuracy

Recognition accuracy is described as the ratio of number of facial images that are correctly recognized to the

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: 25 July 2023 Revised: 12 September 2023 Accepted: 30 September 2023

total number of facial images. It is measured in terms of percentage (%). It is formulated as,

$$RA = \frac{Number\ of\ facial\ images\ that\ are\ correctly\ recognized}{Number\ of\ facial\ images} * 100 \quad (1)$$

From (1), 'RA' represent the recognition accuracy. When the recognition accuracy is higher, the method is said to be more efficient. Table 1describes the recognition accuracy comparison for six different existing methods.

Table 1: Recognition Accuracy Comparison

Number of facial images	Recognition Accuracy (%)					
	BERT model	Automated integrated speech signal and facial image analysis system	End-to-end speech driven facial animation algorithm	Taris Model	Novel multimodal fusion attention network	Bimodal fusion algorithm
50	87	92	85	71	83	79
100	89	94	88	73	86	81
150	92	95	90	75	89	84
200	90	93	87	72	86	82
250	89	91	85	70	81	80
300	86	88	82	68	79	77
350	84	86	80	66	77	75
400	82	84	78	64	75	73
450	80	82	76	62	72	70
500	83	85	79	65	74	72

Table 1 explains the recognition accuracy with respect to number of facial images ranging from 50 to 500. Recognition accuracy comparison takes place on the existing BERT model, automated integrated speech signal and facial image analysis system, End-to-end speech driven facial animation algorithm, Taris Model, Novel multimodal fusion attention network and Bimodal fusion algorithm.

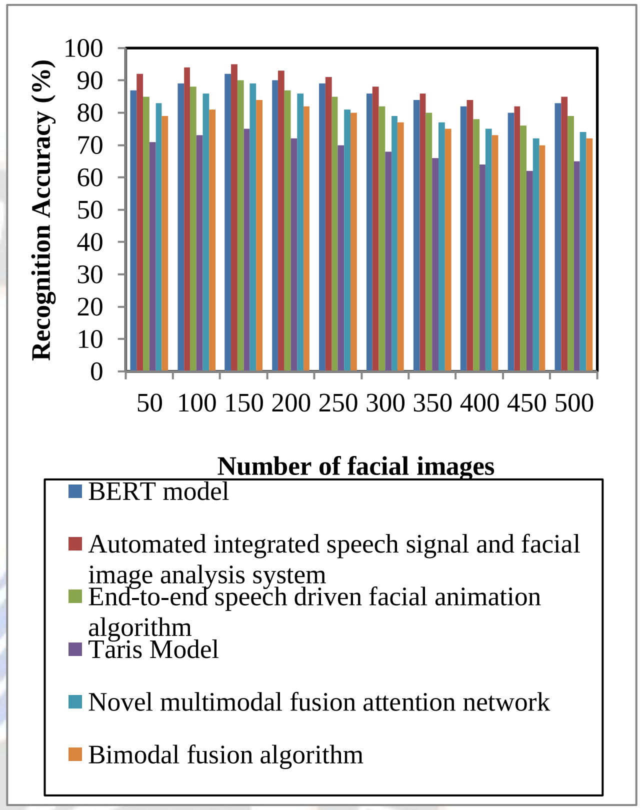

Figure 1: Recognition accuracy with number of face images

The graphical representation of recognition accuracy is shown in the figure 1. Figure 1 illustrates the performance analysis of recognition accuracy with respect to different number of facial images. Let us consider that number of facial images as 300, the recognition accuracy of BERT model, automated integrated speech signal and facial image analysis system, End-to-end speech driven facial animation algorithm, Taris Model, Novel multimodal fusion attention network and Bimodal fusion algorithm is 86%, 88%, 82%, 68%, 79% and 77% respectively. As shown in figure 1, the recognition accuracy using automated integrated speech signal and facial image analysis system is higher when compared to other existing methods. This is because of multi-classification analysis carried out to select the relevant features to recognize the different human emotions. The selected features combination with age and gender are used to construct the learning-based classifiers for achieving higher

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

recognition accuracy. As a result, recognition accuracy of automated integrated speech signal and facial image analysis system is increased by 3%, 7%, 30%, 11% and 15% when compared to BERT model, End-to-end speech driven facial animation algorithm, Taris Model, Novel multimodal fusion attention network and Bimodal fusion algorithm respectively.

5.2 Analysis on Error Rate

Error rate is defined as the ratio of number of facial images that are incorrectly recognized to the total number of facial images. It is measured in terms of percentage (%). It is computed as,

$$ER = \frac{Number\ of\ facial\ images\ that\ are\ incorrectly\ recognized}{Number\ of\ facial\ images} * 100 \quad (2)$$

From (2), 'ER' represent the error rate. When the error rate is lesser, the method is said to be more efficient. Table 2 describes the error rate comparison for six different existing techniques.

Table 2: Error Rate Comparison

Number	Error Rate (%)					
50	35	31	42	18	22	28
100	37	33	45	20	25	30
150	39	36	47	22	27	33
200	36	34	44	19	24	31
250	34	31	42	17	22	29
300	32	29	40	15	20	27
350	34	32	38	18	23	25
400	36	35	41	20	25	28
450	39	37	43	21	28	31
500	41	40	45	23	30	34

Table 2 explains the error rate with respect to number of facial images ranging from 50 to 500. Error rate comparison takes place on the existing BERT model, automated integrated speech signal and facial image analysis system, End-to-end speech driven facial animation algorithm, Taris Model, Novel multimodal fusion attention network and Bimodal fusion algorithm. Let us consider that number of facial images as 400, the error rate of BERT model, automated integrated speech signal and facial image analysis system, End-to-end speech driven facial animation algorithm, Taris Model, Novel multimodal fusion attention network and Bimodal fusion algorithm is 36%, 35%, 41%, 20%, 25% and 28% respectively. The graphical representation of error rate is illustrated in the figure 2. Figure 2 illustrates the performance of error rate for different number of facial images. Above graphical figure indicates that the error rate using Taris Model is reduced when compared to other conventional methods.

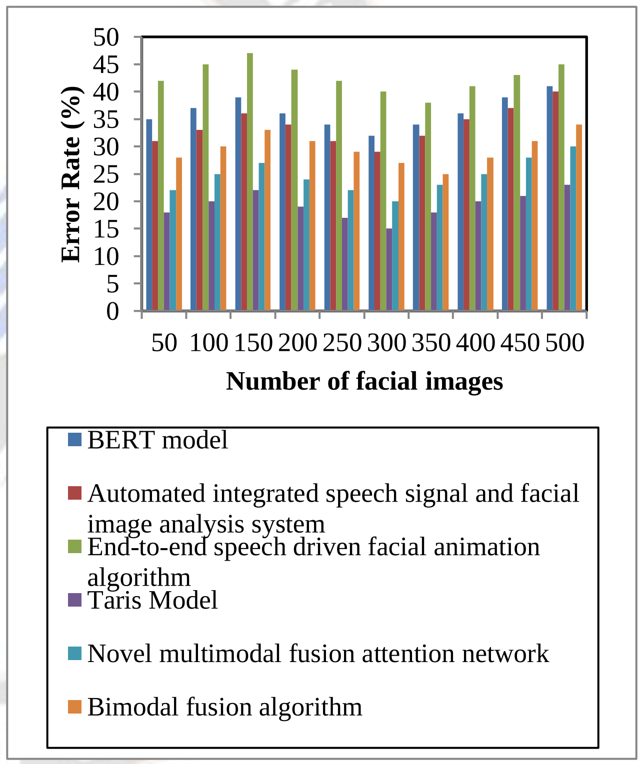

Figure 2: Error rate with number of face images

This is due to the Taris Model has the potential to use the Audio-Visual Speech Recognition (AVSR) technology. A multimodal extension of Taris system obtained the fully differentiable solution to improve the online audio-visual speech recognition and minimize the error rate. As a result, error rate of Taris Model is reduced by 47%, 43%, 55%, 22% and 35% when compared to BERT model, automated integrated speech signal and facial image analysis system, End-to-end

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

speech driven facial animation algorithm, novel multimodal fusion attention network and bimodal fusion algorithm respectively.

5.3 Analysis on Recognition Time

of total number of facial images and time consumed to recognize one face image. It is measured in terms of milliseconds (ms). It is calculated as, Recognition time is defined as the product

$$RT = Number\ of\ images * Time\ consumed\ to\ recognize\ one\ image \quad (3)$$

From (3), 'RT' denotes the recognition time. When the recognition time is lesser, the method is said to be more efficient. Table 3 explains the recognition time comparison for six different existing techniques.

Table 3: Recognition Time Comparison

Number of facial images	Recognition Time (ms)					
	BERT model	**Automated integrated speech signal and facial image analysis system**	**End-to-end speech driven facial animation algorithm**	**Taris Model**	**Novel multimodal fusion attention network**	**Bimodal fusion algorithm**
50	40	35	18	33	12	24
100	42	38	20	35	15	26
150	45	41	23	37	17	29
200	47	43	26	39	20	31
250	49	46	29	41	22	33
300	51	49	32	43	25	35
350	53	51	35	45	28	38
400	55	53	38	48	30	40
450	58	56	40	50	33	42
500	60	58	42	52	35	44

Figure 3 illustrates the performance of recognition time for different number of facial images. The observed graphical results indicate that the performance of recognition time using novel multimodal fusion attention network is minimized when compared to other conventional methods. This is because of applying an adaptive and multi-level factorized bilinear pooling. A global factorized bilinear pooling approach is also applied to combine the audio-visual information by integrating the self-attention based video stream with audio stream.

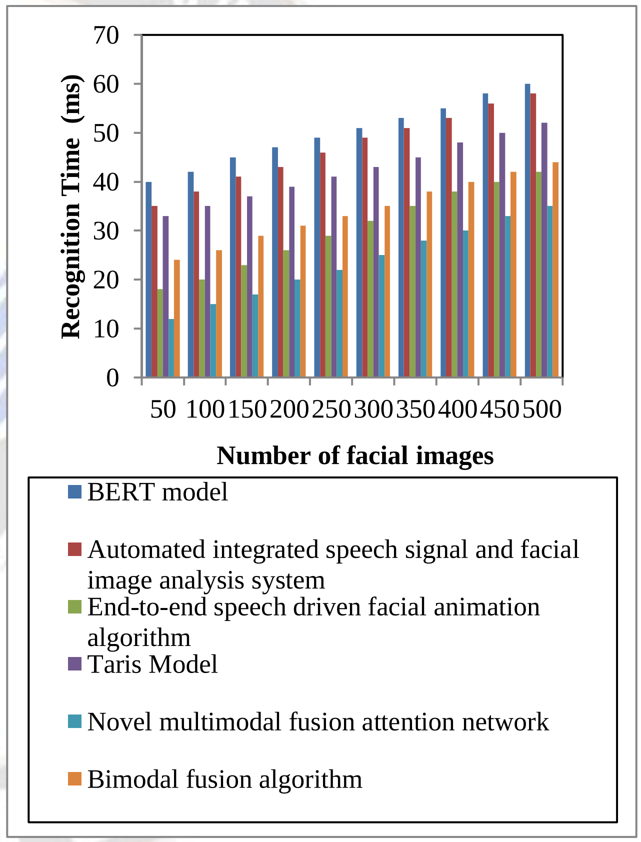

Figure 3: Recognition time with number of face images

This process minimizes the recognition time. As a result, recognition time of multimodal fusion attention network is considerably minimized by 54%, 51%, 23%, 45% and 32% when compared to BERT model, automated integrated speech signal and facial image analysis system, End-to-end speech driven facial animation algorithm, Taris Model and Bimodal fusion algorithm respectively.

6. DISCUSSION AND LIMITATION ON SECURED USER AUTHENTICATION THROUGH BIOMETRIC RECOGNITION

A novel multimodal emotion recognition approach enhanced the performance of BERT model for emotion recognition. The designed approach combined the fine-grained audio and visual features representation with fine-tuning modalities. However, the computation cost of the designed approach was higher due to the trainable weight and hyper parameters. An automated integrated speech signal and facial image analysis system was developed for identification of different human emotions by selecting the relevant features and automatic analysis of free speech and image face analysis. The designed model increased the emotions identification accuracy. However, it was not efficient to find the significant evaluation metrics for detecting the person emotion with mobile phone recordings and images.

An end-to-end speech driven facial animation algorithm developed for pre-trained speech recognition by creating the lip-synchronized facial animation and robust vocal features from vocal audio clip. But, designed algorithm failed to have the ability of controlling the emotion or speaking style for improving the expressiveness of facial animation. A fully differentiable neural network approach termed Taris developed to decode the audio-only and audio-visual speech. But, the time complexity of the decoding process was not minimized.

A novel multimodal fusion attention network developed for audio-visual emotion recognition by using adaptive and multi-level factorized bilinear pooling. The designed network integrates the temporal feature sequence to single label. But, the error rate during the emotion recognition was not minimized. The limitation of this proposed method investment required to biometrics for security; biometric databases can still be hacked as well as biometric devices like facial recognition systems can limit privacy for users.

A bimodal fusion algorithm introduced for speech emotion recognition through the facial expression and speech information. However, the designed algorithm consumed more recognition time.

7. Conclusion

This paper analyzes comparison of various biometric recognition techniques such as BERT, an automated integrated speech signal and facial image analysis system, End-to-end speech driven facial animation algorithm, Taris Model, Novel multimodal fusion attention network and Bimodal fusion algorithm. From the study, it is observed that the recognition accuracy was not improved by bimodal fusion algorithm. A fully differentiable neural network approach called Taris failed to minimize the recognition time complexity. In addition, the error rate of emotion recognition was not reduced by using novel multimodal fusion attention network. The wide range of experiment on existing methods concludes that the performance of biometric recognition techniques with its limitations. Also, the biometric recognition techniques are compared based on various metrics, for example, recognition accuracy, and time and error rate. The experimental results concluded that the machine learning and deep learning techniques for improving the biometric recognition provides high 13% accuracy and 40% of minimum error rate and 41% lesser time consumption.

8. Future Direction

The future direction of study is to perform efficient user authentication through biometric recognition with higher accuracy and lesser error rate as well as time consumption by applying machine learning and deep learning techniques.

REFERENCES

[1] Wang, J.S. Exploring biometric identification in FinTech applications based on the modified TAM. *Financial Innovation, Springer*, 7 (42): 1-24, 2021.

[2] Mansouri-Benssassi, E. and Ye, J. Generalisation and robustness investigation for facial and speech emotion recognition using bio-inspired spiking neural networks *Soft Computing, Springer*, 25: 1717–1730, 2021.

[3] Lee, S., Han, D.K., and Ko, H. Multimodal Emotion Recognition Fusion Analysis Adapting BERT with Heterogeneous Feature Unification. IEEE Access, 9: 94557 – 94572, 2021.

[4] Lavan, N., Smith, H., Jiang, L., and Gettigan, M.C. Explaining face-voice matching decisions: The contribution of mouth movements, stimulus effects and response biases. *Attention, Perception, & Psychophysics, Springer*, 83: 2205–2216, 2021.

[5] Chai, Y., Weng, Y., Wang, L., and Zhou, K. Speech-Driven Facial Animation with Spectral Gathering and Temporal Attention. *Frontiers of Computer Science,* 16: 1-15, 2022.

[6] Shahamiri, S.R. Speech Vision: An End-to-End Deep Learning-Based Dysarthric Automatic Speech Recognition System. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 29:852-861, 2021.

[7] Wang, X., Chen, X., and Cao, C. Human emotion recognition by optimally fusing facial expression and speech feature. *Signal Processing: Image Communication*, 84: 1-6, 2020

[8] Saxena, N., and Varshney, D.Smart Home Security Solutions using Facial Authentication and Speaker Recognition through Artificial Neural Networks. *International Journal of Cognitive Computing in Engineering*, 2: 154-164, 2021

[9] Bansal, V. and Garg, S. A cancelable biometric identification scheme based on bloom filter and format-preserving encryption. *Computer and Information Sciences*, 34: 5810–5821, 2022.

[10] Tse, K-W. and Hung, K. Framework for user behavioural biometric identification using a multimodal scheme with keystroke trajectory feature and recurrent neural network on a mobile platform. *Institute of Engineering And Technology*, 11: 157-170, 2021

[11] Faruk, J.H., Islam, M., Alam, F., Shahriar, H., Rahman, A. BieVote: A Biometric Identification Enabled Blockchain-Based Secure and Transparent Voting Framework. *Department of computer science and Engineering:* 2022.

[12] Pereira, T.M.C., Conceição, R.C., Sencadas, V., and Sebastião, R. Biometric Recognition: A Systematic Review on Electrocardiogram Data Acquisition Methods, Sensors, MDPI, 23 (3): 1-31, 2023.

[13] Almomani, I., El-Shafai, W., AlKhayer, A., Alsumayt, A., Aljameel, S.S., and Alissa, K. Proposed Biometric Security System Based on Deep Learning and Chaos Algorithms. Computers, *Materials & Continua* , 74(2): 3515-3537, 2023

[14] Elazm, L.A.A., El-Shafai, W., Ibrahim, S. Efficient Hardware Design of a Secure Cancellable Biometric Cryptosystem. *Intelligent Automation & Soft Computing*, 36 (1): 929-955, 2023.

[15] Harikrishnan, D., Sunilkumar, N., Shelby, J., Kishor, N., Remya, G. An effective authentication scheme for a secured IRIS recognition system based on a novel encoding technique. Measurement Sensors, 25: 100626, 2023